New York State Economics Association

Proceedings of the New York State Economics Association Conference

Volume 15

75th Annual Meeting, Oct. 7-8, 2022

Editor: Manimoy Paul

Published: October 2022

75th Annual Meeting NYSEA Proceedings, Volume 15

New York State Economic Association

Founded 1948

Proceedings 2022

Published: October 2022

Table of Contents

Venture Capital: Case of Shark Tank.

Shalei Simms[1], Elena Smirnova [2], Zhihong Shi [3], Christopher Castro-Andrade [4]

ABSTRACT

We study small scale venture capital investment through the popular reality show *Shark Tank,* looking at the financial impact of celebrity influence over venture capital allocation. We use a unique hand-collected dataset and apply Becker's theory of discrimination to the *Shark Tank* simulation of the VC market. We find that entrepreneurs are able to learn that it would be costly to have a preference for glamourous celebrity backers. They will not give up higher equity stakes during the funding negotiations with celebrities. Additionally, celebrities' investment does not add value to the projects, even in the era of powerful social media.

INTRODUCTION

Numerous business ideas are financed by venture capital. Investors seeking high rates of return provide funds for new projects through venture capitalists (VCs). In addition to capital formation, venture capitalists offer strategic insight, management expertise, network partners, and possibilities for future funding (Matherne, 2010).

Shark Tank is a TV show where entrepreneurs can pitch their products to the panel of "Shark" investors, seeking funding to expand their businesses, most often single-product ventures. The *Shark Tank* investors include both the regular panelists, who become TV stars in their own right, as well as guests, most often celebrities from the world of sports, entertainment and sometimes even business. The pitching entrepreneurs usually offer some equity in their businesses in exchange for funding. The show is now in its 14th season. It offers an entertaining opportunity for small business to access capital and expertise that would otherwise not be available, on a much sought-after public stage that also offers huge marketing exposure.

Shark Tank is a long running hit show and a pop culture phenomenon. For many Americans (and people around the world), it is their most in depth exposure to the world of business, investment, and startups. What do the key results from *Shark Tank* tell us about the business financing system, and about how *Shark Tank* is shaping the popular perception of it?

[1, 2, 3] Department of Management, Marketing and Finance, School of Business, State University of New York at Old Westbury.

[4] Undergraduate Student in Finance, Department of Management, Marketing and Finance, State University of New York at Old Westbury.

The investors on the panels, including the guest panelist, are not conventional venture capitalists, with Ivy League MBAs and Sandhill Road offices. They are all self-made successful business people with their own entrepreneurial life experiences. *Shark Tank* provides a reality show stage on which investors and entrepreneurs who otherwise might never be able to meet can come together to seed new business ideas and create an engaging mass market spectacle at the same time.

From the finance perspective, entrepreneurs would otherwise find barriers to enter conventional venture capital markets, with its intense competition for funding from investment firms and private equity funds. Hochberg, Ljungqvist and Lu (2007, 2010) find that proximity of VCs to the market is the single predictor of outside VC entry into the new market. *Shark Tank* creates a perfect and extremely proximate marketplace for both young entrepreneurs and potential early-stage investors. Through 13 seasons, $ 169,533,029 of has been raised through the show.

The five regular panelists on the show have become household names – Mark Cuban has talked about running for President, another panelist is known as "Mr. Wonderful," – the inevitable results of starring in a long running, top rated network show. Then there are the "guests," panelists who come on the show for multiple episodes in a season, who are also stars, perhaps of even higher magnitude than the regular panelists. They have included Alex Rodriguez, Gwyneth Paltrow, Richard Branson, Ashton Kutcher and Kevin Hart, among others.

While regular panelists account for most of the "investments" made on the show, the guests often compete to make offers accepted by the fund seekers. In this paper, we seek to explore if the fund seekers are more eager to accept offers from guests than from the regular panelists. While *Shark Tank* is hardly a topic that cries out for considerations of social justice, we apply Gary Becker's theory of discrimination (Becker, 1957) to our analysis of the panelist-guest dynamic.

Our analysis does not explain the motivation behind the fund-seekers' decisions. Perhaps they're simply star struck, and they derive non-pecuniary benefit from an association with a famous investor. Or perhaps the fund-seekers are more calculating, anticipating that association with an investor with wide recognition, millions of social media followers, will be a tangible asset in launching the product.

If the fund-seekers impute added value to celebrities, they would be willing to accept financially less beneficial celebrity offers and would be willing to give up a comparatively higher percentage in equity to celebrity investors. As Becker (1957) suggested, the discriminator and the person being discriminated against (or for) will both experience economic consequences. This will make celebrity guests' investment portfolios enjoy higher rates of return compared to that of the regular sharks' portfolios.

On the other hand, parallel to Becker's (1957) argument about how non-discriminating employers can take advantage of the racial wage differences, we propose that some entrepreneur fund seekers are aware of the premium the celebrity investors are demanding and decide to take the resident Shark offers and to benefit from the business expertise of the more experienced panelist investors.

PRIOR LITERATURE AND HYPOTHESES

Becker's (1957) book *The Economics of Discrimination* revolutionized social science research in discrimination. Lang and Spitzer (2020) provide an excellent review of the labor market and criminal justice system. There are several studies in mortgage lending discrimination (Bauer and Cromwell, 1994; Berkovec, Canner, Gabriel and Hannan, 1998). The theory has also been applied to research not related to discrimination.

Becker's (1957) showed that if a person discriminates against other people on grounds other than productivity, the discriminator will bear the cost of his choice. The person being discriminated against will be hurt, too. In the case of racial discrimination in wages, the black workers will earn less as compared to white workers. The discriminating employer will have to pay more to get white workers with the same productivity.

In *Shark Tank*, entrepreneurs may have a preference to work with celebrities, simply because of the glitter. They are willing to accept lower funding offers and / or give up higher equity shares in exchange for the funds they request. On the other hand, entrepreneurs may value the celebrities' advantage in promoting a project and thus boosting sales, suggesting that the entrepreneurs see a productivity value in celebrity investors. As in Becker's (1957) theory, both sides of the discriminating choice would face economic consequences. The entrepreneurs' discrimination against the main Sharks and the favoritism for the celebrities will help celebrities make higher returns on their investments.

Hypothesis 1: Celebrities would have a competitive advantage in bidding against the resident Sharks. The probability of winning the bid is higher for celebrity guests.

Hypothesis 2: The celebrities would make higher returns on their investments by acquiring higher equity stakes in the projects than the capital requested by the entrepreneurs (Celebrity "premium").

However, market forces could tend to reduce the contestant celebrity preference and thus reduce that premium. Becker (1957) applies to a market and assumes that market participants are rational and can learn to move away from racial hiring bias. There are incentives for non-discriminating employers to hire black workers to take advantage of lower labor cost. If there are enough non-discriminating employers in the marketplace doing the same thing, racial wage discrimination could be eliminated. In the case of *Shark Tank,* entrepreneurs could enhance their decision making by studying past shows. When they realize that celebrities were often able to join the pitched project with lower funding and / or get higher equity stakes, the entrepreneurs would be more willing to consider and eventually take offers from the resident Sharks. Being partners with resident Sharks, the entrepreneurs will give up less of their equity stake and benefit from the real business expertise of the "career investor" Sharks. When there are enough entrepreneurs who choose to partner with the resident Sharks, the celebrity "premium" should disappear.

If *Shark Tank* reflects the learning process of the market participants, the two hypotheses above (regarding entrepreneur celebrity preference) would be rejected.

Both celebrity guests and resident Sharks create value for the funded projects. Celebrities add value through their powerful influence on their fans. Resident Sharks enhance the projects with their business

expertise and networks. In the era of powerful social media, we try to examine if celebrities' far reaching fan base is more powerful than the benefits of business expertise. Since the data on the returns of the projects after funding is not available, we use logistic regressions to study the projects' survival rates based on their investors.

Hypothesis 3: The projects funded by celebrity Sharks would have higher survival rates than those funded by resident Sharks.

DATA AND METHODOLOGY

We collect data from the *Shark Tank* website (www.sharktanktales.com) and websites of individual entrepreneurs. The variables of interest include the amount of funds requested, the amount raised, and structure of a deal (equity stakes, debts, loyalties and promissory notes). Since the returns of the funded projects are not available, we collect data on the survival of the company after it appeared on *Shark Tank*, and how many episodes each Shark appeared in.

Industry distribution of the products on *Shark Tank* from August 9, 2009, until November 19, 2021, is presented in Table 1. Most of the venture capital projects are in the Consumer Discretionary category, accounting for 70.31% of our sample. Consumer Staples are the second largest group with 18.86%. Our sample has 1,108 new products that were pitched on the show.

Table 1. Industry distribution of the products pitched on Shark Tank (Aug. 2009 – Nov. 2021)

Industry	**Number of projects**	**Percentage of total projects**
Consumer discretionary	775	70.31%
Consumer staples	208	18.86%
Communication services	56	5.05%
Utilities	30	2.71%
Healthcare	20	1.81%
Industrials	6	0.54%
Financials	6	0.54%
Real estate	2	0.18%
Total	1,108	100%

After we drop the products that failed to get funding on the show, we have a sample of 651 observations that received some sorts of funding: equity stake, a loan, a royalty or a promissory note. Table 2 presents descriptive statistics for our remaining sample of funded projects.

Table 2. Descriptive statistics for the sample of 651 funded projects (Aug. 2009 – Nov. 2021)

Variable	Mean	Standard deviation	Minimum	Maximum
Funds requested	$260,419.40	$327,972.80	$10,000	$5,000,000
Change in equity stake after the project was funded	12.17%	12.85%	0	95%
Equity percentage	13.61%	8.01%	1%	51%
Number of episodes	27.82	19.7	1	72

It is worth noting that entrepreneurs on average request $260,419.40 from venture capitalists, with a 13.61% average equity stake. On average, each Shark appears in 27.82 episodes. Each episode presents 6 to 8 pitches by entrepreneurs.

There are 36 projects funded by celebrity venture capitalists, which constitutes 5.53% of the sample of funded projects. Joint ventures between celebrities and resident Sharks account for 43 funded projects, or 6.6% of the sample. Resident sharks funded the most pitched products, 444 out of 651 (or 68.2% of the sample). Mark Cuban has the maximum appearance of 72 episodes. Resident alliances of several sharks funded 127 projects in our sample (19.51%). Appendix provides distribution of projects funded by visiting celebrities and is available upon request.

We use the projects' survival rate to measure the added value from the Sharks. We create dummy variables to code the survival of projects after the show. Survival equals one if the project is operational now, zero if it did not survive, and two if the company was merged or a venture was acquired after the show. The number of celebrity Sharks in each episode varies from zero (no resident sharks, just celebrity investor) to 2 on the 5-person panel. We code the gender of the entrepreneur as 1 for male, 0 for female, and 2 for a team of entrepreneurs of different genders. Female entrepreneurs constitute 19% of the sample.

Eighteen projects (2.8% of the funded projects) have been funded by a loan; 46 projects (7% of the funded projects) were funded by royalties; and most of the pitched products were funded by the equity stakes in the future endeavor.

Dependent variable is the probability of a celebrity's win. The entrepreneurs' preference to choose a celebrity as a project partner is reflected in the likelihood of the celebrities' winning a bid. We use a dummy variable of 1 to represent that a project is funded by the celebrity guest Sharks. Independent variable is Celebrity status. Even though every VC is a celebrity because of the popularity of *Shark Tank,* we think some of them are bigger stars than others. We label guests to the show as celebrities. Appendix provides the list of guest celebrities and is available upon request.

The survival of the funded projects is highly correlated with the quality of the projects. Therefore, we incorporate control variables for project quality in the regression. One control variable is the sales volume

of the product at the time the show aired. We also control for the numbers of Sharks in the funded projects. It would be highly beneficial to the projects' performance if Sharks with more industry experiences, capital availability to work with Sharks with better social networks (Hobhberg, Lindsey and Westerfield, 2015).

Based on the discussion above, we have the following hypotheses and corresponding regressions.

Hypothesis 1: Celebrities would have a competitive advantage in bidding against the resident Sharks. The probability of winning the bid is higher for celebrity guests.

$$Probability\ (Celebrity\ win) = \alpha + \beta * Shark_{\ w\ matching\ expertise} + \gamma * Ent\_Gender + \varepsilon \quad (1)$$

Hypothesis 2: The celebrities would make higher returns on their investments by acquiring higher equity stakes in the projects than the capital requested by the entrepreneurs (Celebrity "premium").

$$Increase\ in\ equity\ stake = \alpha + \beta * Survival + \gamma * Investor\ category + \varepsilon \quad (2)$$

Hypothesis 3: The projects funded by celebrity Sharks would have higher survival rates than those funded by resident Sharks. We use the increase in equity stake acquired as the proxy for the quality of the project.

$$Survival\ category = \alpha + \beta_1\ Change\ in\ Equity + \beta_2\ Number\ of\ Sharks + \beta_3\ Ent\_Gender + \varepsilon \quad (3)$$

CELEBRITY STATUS AND VENTURE MARKET

The goal of this paper is to gauge the significance of celebrity presence in the VC market. We propose that quality of the project, celebrity status of the VC, and cooperation of the sharks in the deal affect the change in equity stake of the funded products. We hypothesize that celebrity status of the VC may or may not create a competitive advantage in the bidding against the conventional venture capitalist with expertise in the industries.

We start with Hypothesis 2 of the presence of the Celebrity premium. The objective is to find out if celebrity VCs obtain higher equity stakes in the project seeking funding. Following Backer (1957), if the VC market prefers celebrity VCs, guest celebrity Sharks would be able to acquire higher equity stakes for the funds requested by the entrepreneurs and achieve higher return on their investments. Also, in Becker's model, the market can learn that it is costly to have a bias not based on productivity. If entrepreneurs learn that it is costly to accept celebrities' offers, celebrity status will lose its attractiveness.

We run an OLS regression on the *Increase in equity stake* in the funded project as a function of project survival after the episode aired, and investor category. The survival of the project is coded as a dummy variable that takes on a value of 1 if the funded project survived after the show aired, and the product is selling on the market at present. It takes the value of 0 if the product ceased to exist after it was funded, and the value of 2 if the company merged or the project was acquired by another entity. Investor category is coded as a categorical variable, with 4 types of investors: celebrity guests (category 1), celebrity and resident shark team (category 2), single resident shark (category 3), and multiple resident sharks (category 4).

Panel (1) of Table 3 shows that survival of the projects and investor category play a significant role in the increases in equity stakes of the funded projects. Survival of the projects after the show are negatively

related to increases in equity stakes in the venture. We interpret this finding as evidence of higher success rates for projects with less equity stake offered to the venture capitalist. The owners strive to maintain control of the products and their development after the show, while using venture capital from the investors.

Investor category is an indicator variable which goes from 1 (celebrity Shark only) to 4 (resident Sharks only) depending on the type of venture capitalist financing the deal. A positive coefficient of investor category in Panel (1) of Table 3 means that as investor category changes away from celebrity investors to resident Sharks, the equity percentage goes up. This finding is an indication of a negative celebrity premium, where entrepreneurs offer less equity stake to celebrities. This result shows that the market learns.

Panel (2) of Table 3 shows the linear regression model with celebrity investor as one of the categories of interest, and number of sharks in a deal.

$$\textit{Increase in equity stake} = \alpha + \beta_1 \textit{Survival} + \beta_2 \textit{Celebrity} + \beta_3 \textit{Number of Sharks} + \beta_4 \textit{Sales} + \varepsilon \quad (4)$$

The coefficient of interest is β_2. If entrepreneurs have biased preferences for celebrity Sharks, we will find positive coefficient on celebrity status of the venture capitalist, indicating that the entrepreneur is offering a higher equity stake in the project to celebrity investors, and also indicating an implied celebrity premium. If entrepreneurs are able to learn that the celebrity premium is costly, they would not give up too much equity stake. Our results support the market learning model - the coefficient is negative and statistically significant. Based on our sample, celebrity-financed projects have a 3.23% less equity stake increase compared to projects financed by non-celebrity investors.

Table 3. Linear regression results.

	Increase in Equity percentage (Panel 1)	**Increase in Equity percentage (Panel 2)**
Constant	0.0804***	0.0978***
Survival	-0.0337**	-0.0326**
Investor Category	0.0235***	
Celebrity investor		-0.0323**
Number of sharks in deal		0.0428***
R-squared	0.0301	0.057
Number of observations	651	651

*** denotes significance at 1% level, ** at 5% level and * at 10% level.

We examine whether the odds of projects survival are affected by percentage of equity financing, number of sharks participating in the project financing, as well as entrepreneurs' gender. We coded survival categories as 0 if the funded product ceased to exist after the show; 1 for the funded product being on the market after the show; and 2 for the project being merged or acquired by another entity after it was funded.

$$Odds\ of\ Survival = \alpha + \beta_1\ Change\ in\ Equity + \beta_2\ Number\ of\ Sharks + \beta_3\ Ent\_Gender + \varepsilon \quad (5)$$

Table 4, panel (1), reports the multinomial logistic regression results with relative risk ratios (RRR) and their corresponding z-scores for the survival of the project relative to non-survival, as explained by change in equity percentage, number of sharks in the project and gender of entrepreneur. RRR shows the direction, but not the magnitude, of the effect. RRR of less than 1 indicates that the variable has a negative effect on survival relative to non-survival, and the z-score is negative. A relative risk ratio above 1 shows positive influence of an explanatory variable on survival. It is noteworthy that equity stake change is negatively related to odds of survival. The higher the financing stake that entrepreneur gives up in a project, the less is the survival probability relative to non-survival.

Panel (2) of Table (4) reports the results of equation (3) for the merger or acquisition odds relative to non-survival. The variables are statistically insignificant in that specification. This means that funded projects are equally likely to fail or to get acquired after the show aired. It suggests that joint venture of the VCs may not help to create value in the relatively small projects presented in *Shark Tank.*

Table 4. Multinomial Logistic Regression results.

	Outcome: Survival of the project. Relative risk ratios. (Panel 1)	**Z value**	**Outcome: Merger or Acquisition of the project. Relative Risk Ratios. (Panel 2)**	**Z value**
Constant	9.606***	7.16	0.1656**	-2.45
Change in Equity	0.1109***	-2.98	0.1565	-0.95
Number of sharks in a project	0.8437	-1.06	1.0809	0.23
Gender of entrepreneur	0.9371	-0.26	1.4128	0.59

Pseudo McFadden	0.0169
R2	
N	651

Note: Base outcome is Survival = 0 (project is not operating as of 2022). Relative risk ratio (calculated as Exponent to the power of the regression coefficient) of less than 1 means that as the explanatory variable increases, the likelihood of survival (or merger) goes down relative to base category of 0 survival. That is confirmed by negative z scores for RRR<1. RRR higher than 1 means that survival is more likely with an increase in explanatory variable.

CONCLUSIONS AND FUTURE RESEARCH

This paper attempts to estimate the significance of celebrity status in venture capital funding using the show *Shark Tank*. We claim to see parallels in Becker (1957) discrimination theory between two groups of investors: celebrity and non-celebrity VCs. We find significant differences between these two VC types. The relation to the change in equity funding stakes lends support to Becker's market learning hypothesis. Using the sample of 651 funded projects we find that celebrities who fund the products on Shark Tank are not able to enjoy higher equity financing perks given by entrepreneurs since the entrepreneurs are aware of the celebrity premium. Furthermore, a higher change in equity stake of the funded project is related to lower survival rate of the project on average. These findings indicate that offering a lower equity stake is beneficial for the project survival after the show. The celebrity status of the VC is not beneficial in the equity stake bargaining process.

Further research involves collecting data points on the caliber of the entrepreneur, including education and prior industry experience. In addition, we would like to calculate the celebrity premium in terms of rates of return on the projects. Celebrity guests' investment portfolios might or might not enjoy higher returns than those in the portfolios of regular Sharks ("Celebrity premium"). The returns on the celebrities' portfolio are the monthly rate of return on the projects after the show aired. We plan to collect the return data for at least 5 years after the episode aired.

This research has various implications for academics and practitioners alike. First, we can show that giving up a controlling stake in venture capital financing leads to lower survival rates of the venture. This implies that entrepreneurs are better off retaining more control of the company. The second implication is for celebrity status of the venture capitalist: it is not sufficient for the entrepreneur to have the glamor of the celebrity VC since the equity stake is lower.

REFERENCES

Bates, T., Bradford, W. D. and William E. Jackson III (2018). Are minority-owned businesses underserved by financial markets? Evidence from the private equity industry. Small Business Economics, 50, 445-461.

Bauer, Paul, and Brian Cromwell. "A Monte Carlo Examination of Bias Tests in Mortgage Lending," *Federal Reserve Bank of Cleveland, Economic Review,* vol. 30, no. 3, pp. 27-40, 08.01.1994.

Becker, Gary (1957). The economics of discrimination. University of Chicago Press.

Berkovec, J.A., Canner, G.B., Gabriel, S.A. *et al.* (1994). Race, redlining, and residential mortgage loan performance. *Journal of Real Estate Financial Economics* 9, 263–294. https://doi.org/10.1007/BF01099279

Brander, J. A., Amit, R., & Antweiler, W. (2002). Venture-capital syndication: Improved venture selection vs. the value-added hypothesis. *Journal of Economics & Management Strategy*, *11*(3), 423-452.

Bone, S., Christensen, G., and Williams, J. (2014). Rejected, shackled and alone: the impact of systematic restricted consumer choice among minority customers' construction of self. Journal of Consumer Research, 41, 451-474.

Gompers, Paul A., and Josh Lerner, 2000a, Money chasing deals? The impact of fund inflows on private equity valuations, Journal of Financial Economics 55, 281-325.

Hochberg, Y. V., Lindsey, L. A., & Westerfield, M. M. (2015). Resource accumulation through economic ties: Evidence from venture capital. *Journal of Financial Economics*, *118*(2), 245-267.

Hochberg, Y. V., Ljungqvist, A. and Lu, Y. (2010). Networking as a barrier to entry and the competitive supply of venture capital. Journal of Finance, 65 (3), 829-859.

Hochberg, Y. V., Ljungqvist, A. and Lu, Y. (2007). Whom you know matters: Venture Capital networks and Investment Performance. Journal of Finance, 62 (1), 251-301. https://www.jstor.org/stable/4123462

Lerner, J. (1994). Venture capitalists and the decision to go public. *Journal of financial Economics*, *35*(3), 293-316.

Lang, Kevin, and Ariella Kahn-Lang Spitzer. 2020. "Race Discrimination: An Economic Perspective." *Journal of Economic Perspectives*, 34 (2): 68-89.

Longhofer, Stanley D. (1996). Discrimination in Mortgage Lending: What Have We Learned?

Matherne, B. P. (2010). Local VC Networks: Barriers to Entry for New Venture Capital? Research Brief. Loyola University, New Orleans.

Munnell, A. H., Tootell, G., Browne, L and McEneaney (1996). Mortgage lending in Boston: Interpreting HMDA data. *The American Economic Review* (1996): 25-53.

Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American Journal of Sociology*, 106 (6), 1546-1588.

Sorenson, O., & Stuart, T. E. (2008). Bringing the context back in: Settings and the search for syndicate partners in venture capital investment networks. *Administrative Science Quarterly*, *53*(2), 266-294.

Souitaris, V., Zerbinati, S. and Liu, G. (2012). Which iron cage? Endo- and exo- isomorphism in corporate venture capital programs. Academy of Management Journal 55 (2), 477-505. http://dx.doi.org/10.5465/amj.2009.0709

Zhang, L., Gupta, A. Hallen, B. (2017). The conditional importance of prior ties: A group-level analysis of venture capital syndication. Academy of Management Journal, 60 (4), 1360-1386. https://doi.org/10.5465/amj.2013.1088

The Causal Effect of Health on Out-of-Pocket Health Expenditures: Evidence for Middle-aged and Older Adults in China

Jingyi Gao[5]

ABSTRACT

With population aging, the presence of disability and chronic conditions has become a growing challenge among middle-aged and older adults in China. Using panel data of China Health and Retirement Longitudinal Study (CHARLS) from 2011 to 2018, this paper will explore the causal effects of Activities of Daily Living (ADL) limitations, Instrumental Activities of Daily Living (IADL) limitations, chronic conditions on Out-of-Pocket Expenditures (OOPEs) among middle-aged and older adults in China. First difference model and difference-generalized method of moment (GMM) model are used to explore the causal relationships. This research for the first time examines the dynamic and causal relationship between health and OOPEs.

INTRODUCTION

The share of older adults in the broader population in China has been rapidly growing. the older population (aged 60 years or older) is reaching 18,7% of the population (National Bureau of Statistics, 2021). High prevalence of disability and chronic conditions are associated with aging. For instance, around 75% of the older population are experiencing chronic conditions including diabetes, hypertension and other chronic conditions (WHO, 2019). It is found that 26.2% of the older population in China suffer from disability based on the meta-analysis of 97 studies from 1979 to 2022 (Zheng, 2022). The high prevalence of disability and chronic conditions among the old presents challenges to the public health system in China.

An integrated public health system, which is characterized by the utilization of health services and availability of health insurance and equal access to health care regardless of geographical area tackles the challenges from high prevalence of disability and chronic conditions (Cao, 2011). Aiming to provide universal financial protection to individuals and achieve universal health coverage, the Chinese government has established a series of social health insurance schemes targeting both urban residents in municipal level and rural residents in country level since 1998: the Urban Employee Basic Medical Insurance (UEBMI),

[5]Department of Economics, Fordham University, Bronx, NY 10458.

Urban Resident Basic Medical Insurance (URBMI), and the New Rural Cooperative Medical Scheme (NRCMS) (Dong et al, 2021). In April 2009, the government established another round of comprehensive health system reform involving medical care services policy to provide a universal healthcare system (CPC, 2009). OOPEs have decreased from 56% in 2003 to 29% in 2017 as the social health insurance coverage expands. There has also been a greater health service utilization (Fang, 2019). However, the problems of inequalities of healthcare and insufficiency of cost coverage remain (Xian et al, 2019). Thus, it is imperative to provide reliable analysis to raise public awareness of health issues, inform policymakers to allocate recourses to relieve disparities and improve health care system.

There has been an extensive literature examining the association between higher out-of-pocket expenditures (OOPEs) and health conditions. However, previous studies do not address unobserved heterogeneity and reverse causality from OOPEs to health.

This research paper fills this gap by examining the causal links between health measured by activity limitations and chronic conditions and OOPEs among the middle-aged and older adults in China using panel data of CHARLS from 2011 to 2018. More specifically, this paper will address the research question: do health problems lead to higher OOPEs?

The following sections review the relevant literature, describe the dataset, measures and models and the results. The last section concludes.

LITERATURE REVIEW

Many studies have estimated health care expenditure and extra cost of disability across demographic and socioeconomics groups. Owens (2008) analyze the gender differences in health care expenditures and resource utilizations. He finds that health care expenditures tend to be higher among women than men. The greatest disparity in health care spending between men and women occurred in the population aged 45 to 64 years. Loyalka et al (2014) measure the extra cost of disability across different types of disability and different types of households in both urban and rural areas of China. They find the extra costs of disability are larger for urban households than rural households and there is a strong negative correlation between disability and household income.

There have been many studies to examine the relationship between OOPEs and health problems. They use cross sectional data and estimate a correlation. Zhao et al (2021) use quantile regressions to identify the positive association between multimorbidity and OOPEs in China using CHARLS 2015 data. You and Kobayashi (2011) use 2004 China Health and Nutrition Survey data and apply Heckman selection model and find that individuals who had chronic conditions, earned higher income, resided in urban areas, lived in the middle or eastern region, or lived in a household with a head having a middle school or higher education incurred more OOPEs. They conclude that the perceived severity of illness and self-reported health status are the most important determinants of OOPEs. Salinas-Rodriguez et al (2020) find a positive association between activities of daily living/instrumental activities of daily living dependence and OOPEs using two-part regression model and quantile regression among the Mexican adults aged 60 and older. Nguyen et al (2021) also apply the two-part regression models to find that disability is strongly associated with higher OOPEs in Vietnam. These studies focus on the associations between higher OOPEs and health conditions, but do not address the causal effect of health on OOPEs.

A growing number of studies assess the causal effects of health on economic indicators using dynamic panel data model to overcome limitations of the static models and address reverse causality while assessing underlying two-way causal links. However, these studies do not use OOPES as outcome of interest. Kim and Mitra (2022) examine the two-way causal relationships between health and labor income among the middle-aged and older Koreans. They apply the dynamic panel data model with 12 waves of data from 2006 to 2017. They also stratify the sample based on age, gender, region, income level, and marital status to analyze the outcomes across different demographic and socioeconomic groups. Meraya et al (2017) examine the dynamic relationships between economic status and health measures using 8 waves of panel data. A causal link from labor income to self-rated health and functional status for both genders is found using system-generalized method of moment. Our study follows the dynamic panel data model framework addressing the unobserved heterogeneity and reverse causality from OOPEs to health and for the first time, identify the causal links between health problems, on the one hand, and OOPEs, on the other, among the middle-aged and older adults in China using panel data from four waves of CHARLS from 2011 to 2018.

METHODS

This paper uses the 2011, 2013, 2015 and 2018 waves of the China Health And Retirement Survey (CHARLS), which is representative of both the rural and urban population in China. The baseline survey for CHARLS was conducted in 2011/2012 and included 10,257 households and 17,500 individual respondents. The survey collects information on the demographics, family, health status, health care, health insurance and wealth at both individual and household levels. Specifically, CHARLS has a series of questions on OOPE. All respondents are asked to report how many times they have received outpatient care and inpatient care in the past, the total medical cost of all the doctor visits and the hospitalization, the total cost of doctor visits and hospitalizations. This study focuses on 7,448 respondents aged 50 and over that can be followed over the four waves from 2011 to 2018, which forms a balanced panel data after removing missing values.

Annual OOPEs are the total direct payments for both outpatient and inpatient visits after deducting the reimbursed expenses for the year until the survey. Outpatient OOPEs is measured by the value of out-of-pocket doctors visit expenditures for the respondent in the last month covering the fees paid for treatment, medication costs and prescription drugs. Inpatient OOPEs is measured by the value of out-of-pocket hospitalization expenditure for the respondent in the past year. As out-of-pocket doctor visit expenditure in CHARLS is measured for last month only, we multiplied the value by 12 to estimate annual outpatient expenditures. OOPEs, our target of interest, is thus the sum of outpatient OOPEs and inpatient OOPEs.

Health measures cover disability and chronic health conditions. Disability is measured by Activity of Daily Living (ADL) and Instrumental Activity of Daily Living (IADL) limitations. For ADL limitations, each respondent is asked whether they experience some difficulty performing the designated six tasks including bathing, dressing, eating, getting in/out of bed, using the toilet, and controlling urination. For IADL limitations, each respondent is asked whether they experience some difficulty performing the designated five tasks including using the phone, managing money, taking medications, shopping for groceries, and preparing hot meals. A zero indicates that the respondent did not report any problem with each activity. A one indicates that the respondent reported some difficulty or could not do each activity. ADL/IADL limitation number is created by summing the number of answered questions. ADL/IADL limitation index is constructed by normalizing it to 100.

In addition, we use a measure of chronic health conditions that captures 15 chronic conditions (hypertension, diabetes, cancer, chronic lung disease, heart disease, stroke, psychiatric problem, arthritis, dyslipidemia, liver disease, kidney disease, digestive disease, asthma, depression and memory problem). Each respondent answered the question regarding whether or not they have been diagnosed by the doctor for any of the conditions above. We then count the number of diagnosed chronic conditions for each respondent. A Non-Communicable Disease (NCD) index was constructed by normalizing the count to 100. Multimorbidity is defined as the presence of two or more chronic diseases and is measured by a binary variable coded as 1 for respondents with multimorbidity and 0 otherwise.

A difference between ADL/IADL index and NCD index should be noted that ADL/IADL index is constructed by individual self-report and NCD index is constructed by self-report from doctor diagnosis. Doctor diagnosis will potentially incur larger OOPEs while seeing a doctor and use health service compared with purely self-report activity.

My objective is to identify the causal effect of health on OOPEs and I do this in several steps. I start with a first difference model to remove unobserved time-invariant individual's characteristics (Wooldridge, 2006). The model is specified as follows:

$$\Delta Y_{it} = \beta_1 \Delta H_{it} + \sum_{j=2}^{R} \beta_j \Delta X_{jit} + \Delta \mu_{it} \quad (1)$$

where i indexes individuals and t indexes time periods. ΔY_{it} is the change in the log transformed of Y_{it}, which is individual's OOPEs in US dollars at t. ΔH_{it} is the change in health for individual i between time t and time t-1. H_{it} is measured in turn through an ADL Index, IADL Index and NCD Index at t. β_1 is the main coefficient of interest. ΔX_s represents the change in the time-varying control variables at the individual level including marriage, hukou/living arrangement, household size, moving status (outside the original community). The term $\Delta \mu_{it}$ is the change in the individual specific error term that accounts for the change in time-varying unobservables for the respondent.

The first difference model captures the extent to which changes in OOPEs are associated with changes in health. While it solves the problem of unobserved time-invariant heterogeneity, it will not address the omitted variables bias due to the presence of time-varying unobservables. Moreover, it will not provide a consistent estimate in the presence of reverse causality from OOPE to health. It also does not address the

time-dependence of OOPEs. As current OOPEs may be influenced by past OOPEs, current health status and other variables, our second step is to model the causal relationship from health to OOPEs as follows:

$$Y_{it} = \beta_0 Y_{it-1} + \beta_1 H_{it} + \sum_{j=2}^{R} \beta_j X_{jit} + \alpha_i + \mu_{it} \quad (2)$$

where α_i is individual fixed effect: a dummy variable which takes the value one for respondent i and zero otherwise.

Yet in this model (equation (2)), the individual specific fixed effects may be correlated with explanatory variables. Endogeneity of health measures and OOPEs may cause the estimator to be biased and inconsistent. To address this issue, I follow Arellano and Bond (1991) and take the first difference for equation (2):

$$\Delta Y_{it} = \beta_0 \Delta Y_{it-1} + \beta_1 \Delta H_{it} + \sum_{j=2}^{R} \beta_j \Delta X_{jit} + \Delta \mu_{it} \quad (3)$$

Under Arellano and Bond (1991), the first-differenced GMM model controls for unobserved heterogeneity and provides a consistent estimate even in the presence of reverse causality. The lags of the dependent variable are used as instruments to address the limitations of endogeneity between current and past OOPEs and health.

RESULTS

The descriptive statistics are analyzed on 7,448 respondents aged 50 and over. Table 1 presents the mean OOPEs in dollar and the presence of any OOPEs stratified by groups of demographics including gender, income and Hukou/residence arrangements. The average OOPEs is 396.51 USD among men, while the average OOPEs is 461.66 USD among the women. Women experience both higher occurrence of OOPE and higher OOPEs. Our sample is also stratified into four income groups. The average OOPEs among the high-income and low-income is higher than that among the middle-income group based on our statistical results. Moreover, the higher average OOPEs appear in the groups with urban Hukou and the lower average OOPEs appear in the groups with rural Hukou, where Hukou is a household registration system and categorizing each Chinese citizen as either an agricultural (rural) Hukou holder or a non-agricultural (urban) Hukou holder.

Table 2 displays the presence of ADL limitation, IADL limitation and multimorbidity by sociodemographic characteristics of the respondents. As shown from the table, the presence of disability

measured by ADL limitation and IADL limitation increases as age group increases, while the presence of multimorbidity is relatively stable across the age groups. Moreover, the presence of disability is lower in the above median income group compared to the below median income group, while this disparity is not detected for multimorbidity. The prevalence of disability is lowest in the urban Hukou/urban residence arrangement. We can also conclude that the presence of ADL limitation/IADL limitation/multimorbidity is higher among women and individuals who are not married.

Table 1: Out-of-pocket health expenditures, stratified by gender, income, and Hukou/residence arrangement

Group	Features	OOPEs in USD	Presence of OOPEs	Sample size
Gender	Male	396.51	0.251	3,653
	Female	461.66	0.281	3,795
Income	Low	454.02	0.277	1,867
	Low-middle	385.75	0.279	1,879
	Middle-high	411.50	0.257	1,840
	High	467.68	0.253	1,862
Hukou/Residence	Rural Hukou/Rural Residence	399.17	0.260	4,648
	Rural Hukou/Urban Residence	386.45	0.269	1,379
	Urban Hukou/Urban Residence	597.05	0.283	1,080

Urban Hukou/Rural Residence	505.55	0.322	226	

Table 2: The presence of ADL limitation/IADL limitation/Multimorbidity by sociodemographic characteristics

	Full Sample	**ADL Limitation**	**IADL Limitation**	**Multimorbidity**
Age (year)				
50 - 59	0.503	0.153	0.189	0.530
60 - 69	0.360	0.224	0.260	0.608
70 and above	0.137	0.332	0.391	0.592
Gender				
Male	0.490	0.163	0.187	0.513
Female	0.510	0.242	0.295	0.618
Marital status				
Married and partnered	0.854	0.188	0.226	0.556
Unmarried and others	0.146	0.288	0.337	0.618
Education				
Illiterate	0.501	0.251	0.314	0.592
Primary school	0.242	0.128	0.136	0.515
Secondary school	0.223	0.190	0.215	0.560
College and above	0.033	0.118	0.114	0.600

Income				
Low	0.251	0.282	0.288	0.590
Low-middle	0.252	0.282	0.335	0.613
Middle-high	0.247	0.180	0.218	0.540
High	0.250	0.117	0.127	0.522
Hukou/Residence				
Rural Hukou/Rural Residence	0.624	0.223	0.268	0.561
Rural Hukou/Urban Residence	0.185	0.185	0.224	0.553
Urban Hukou/Urban Residence	0.145	0.144	0.171	0.598
Urban Hukou/Rural Residence	0.030	0.209	0.197	0.596
Observations	7,448	1,512	1,804	4,220

The results of first difference regression (equation (1)) are summarized in Tables 3-1 and 3-2 by subgroups. A positive relationship between health, whatever the measure, and the natural logarithm of OOPEs is found. It can be concluded that the association is larger between health and OOPEs for high-income group compared to low-income group. For instance, the estimation results suggest that for high-income group, a ten unit increase in NCD index raises OOPEs by 63%, while a ten unit increase in NCD index raises OOPEs by 36% for low-income group. Moreover, the largest association between health and association exist in the group of population with urban Hukou.

The results of the first-differenced GMM are summarized in table 4-1 and table 4-2. We find a statistically significant positive causal relationship between health measures and OOPEs. To assess the potential heterogeneity of the results across the demographic groups, our sample is stratified by gender and income. A 10 unit increase in ADL index leads OOPEs to rise by 29% for women, while a 10 unit increase in ADL index leads OOPEs to rise by 19% for men. A 10 unit increase in IADL index leads OOPEs to rise by 15% for women, while a 10 unit increase in IADL index leads OOPEs to rise by 21% for men. It is found that the increase in NCD index leads to higher increase in OOPEs compared to ADL/IADL index among both men and women. Next, we stratify the sample by income level. The same unit increase in health measures results in larger increase in OOPEs for the high-income group than that of low-income group. For instance, a 10 unit increase in ADL index leads OOPEs to rise by 46% among the high-income group, whereas a 10 unit increase in IADL index leads OOPEs to rise by 27% among the low-income group. Similarly, the increase in NCD index leads to higher increase in OOPEs compared to ADL/IADL index among both the high-income group and the low-income group.

We further stratify the sample into four Hukou/Residence arrangements (rural Hukou/rural residence, rural Hukou/urban residence, urban Hukou/urban residence, urban Hukou/rural residence). It is found that the same unit increase in health measures leads to the largest increase in OOPEs in urban Hukou/urban residence arrangement compared to other arrangements. The estimated results are not significant for urban Hukou/rural residence arrangement since the sample size is quite limited.

Table 3-1: First difference regression of natural logarithm of OOPEs, stratified by gender and income

	Male	**Female**	**Low Income**	**High Income**
VARIABLES	(1)	(2)	(3)	(4)
ADL index	0.019***	0.022***	0.019***	0.021***
	(0.002)	(0.002)	(0.003)	(0.004)
IADL index	0.017***	0.015***	0.015***	0.023***
	(0.002)	(0.002)	(0.002)	(0.004)
NCD index	0.060***	0.037***	0.036***	0.063***

	(0.005)	(0.005)	(0.007)	(0.004)
Observations	11,072	11,116	5,508	5,544

*** p<0.01, ** p<0.05, * p<0.1

Table 3-2: First difference regression of natural logarithm of OOPEs, stratified by Hukou/residence

	Rural Hukou/Rural Residence	**Rural Hukou/Urban Residence**	**Urban Hukou/Urban Residence**	**Urban Hukou/Rural Residence**
VARIABLES	(5)	(6)	(7)	(8)
ADL index	0.016***	0.019***	0.025***	0.018*
	(0.002)	(0.002)	(0.003)	(0.004)
IADL index	0.012***	0.020***	0.027***	0.011
	(0.002)	(0.002)	(0.002)	(0.004)
NCD index	0.041***	0.040***	0.060***	0.099***
	(0.005)	(0.005)	(0.007)	(0.004)
Observations	18,840	5,620	4,420	720

*** p<0.01, ** p<0.05, * p<0.1

Table 4-1: Arellano-Bond estimates on natural logarithm of OOPEs, stratified by gender and income

	Male	**Female**	**Low Income**	**High Income**
VARIABLES	(1)	(2)	(3)	(4)

ADL index	0.019***	0.029***	0.027***	0.046***
	(0.002)	(0.002)	(0.003)	(0.004)
IADL index	0.015	0.021***	0.017***	0.018
	(0.006)	(0.002)	(0.002)	(0.012)
NCD index	0.089***	0.090***	0.111 ***	0.140***
	(0.023)	(0.005)	(0.007)	(0.042)
Observations	5,536	5,558	2,754	2,772

** p<0.01, ** p<0.05, * p<0.1

Table 4-2: Arellano-Bond estimates on natural logarithm of OOPEs, stratified by Hukou/residence

	Rural Hukou/Rural Residence	**Rural Hukou/Urban Residence**	**Urban Hukou/Urban Residence**	**Urban Hukou/Rural Residence**
VARIABLES	(5)	(6)	(7)	(8)
ADL index	0.022***	0.018*	0.057***	0.015
	(0.002)	(0.002)	(0.003)	(0.004)
IADL index	0.021***	0.024***	0.038***	-0.013
	(0.002)	(0.002)	(0.002)	(0.004)
NCD index	0.094***	0.048**	0.096**	0.013
	(0.005)	(0.005)	(0.007)	(0.004)
Observations	18,840	5,620	4,420	720

*** p<0.01, ** p<0.05, * p<0.1

DISCUSSION AND CONCLUSION

A positive relationship between health problems and OOPEs has been found from both first difference and Arellano-Bond estimates. This result is consistent with earlier studies on the associations between health and OOPEs (Zhao et al, 2021; Salinas-Rodriguez et al, 2020; Nguyen et al, 2021). Another finding is that the increase in NCD index leads to a larger increase in OOPEs compared to ADL/IADL index. This

result is in accordance with our hypothesis that ADL/IADL index is constructed by individual self-report and NCD index is constructed by self-report from doctor diagnosis. Self-reported index and doctor diagnosed index give different results since individuals are using and paying for health service when diagnosed with chronic diseases, but this is not the case for self-reported disability.

We stratify our sample by subgroups and generate heterogeneous results across the demographic groups. It is found from the descriptive statistics that average OOPEs is higher for female, high income and Urban Hukou/Urban Residence subgroups. Based on the model estimations, the increase in health measures leads to higher OOPEs among high-income group and urban Hukou/urban residence arrangement. These results are consistent with earlier studies (Owen, 2008; You and Kobayashi, 2011; Loyalka et al, 2014). Women tend to use significantly more services and spend more health care dollars than men (Owen, 2008). You and Kobayashi (2011) find that individuals who earned higher income and resided in urban areas incurred more OOPEs. Loyalka et al (2014) also talk about the urban-rural disparity in the extra costs of disability in China.

However, our study has several limitations. First, GMM estimator uses lagged differences as instruments for the level model and lagged levels as instruments for the first-difference model. Only four waves of panel data are included in our study, which provides limited choices of instrument variables. Second, robustness checks have not yet been done in this study. For the future work, I will show whether the instruments used for obtaining the preferred specification are strong and whether the underlying assumptions necessary for obtaining consistent estimates from the Arellano-Bond specification are valid. Third, since people who use health care more and have more OOPEs, maybe more aware of health conditions or diagnosed with health-related problems, current health might be influenced by past health. More future work will be on the specification of potential reverse causality model.

This study makes several contributions. It is the first study to examine the causal links between ADL/IADL limitations, chronic conditions and OOPEs in China. Positive causal effects from health to OOPEs have been identified. Second, this paper assesses the heterogeneity of the results across the demographic and social economic groups. These findings can be used to inform policymakers to improve medical resources allocation and ensure better financial protection towards those living with a disability and chronic diseases.

REFERENCES

Arellano M, Bond S. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies.* 1991;58(2):277–297. [Google Scholar]

Dong W, Zwi AB, Bai R, Shen C, Gao J. Benefit of China's Social Health Insurance Schemes: Trend Analysis and Associated Factors Since Health Reform. *Int J Environ Res Public Health*. 2021;18(11):5672. Published 2021 May 25. doi:10.3390/ijerph18115672

Fang H., Eggleston K., Hanson K., Wu M. China's Health System Reforms: Review of 10 Years of Progress: Enhancing Financial Protection under China's Social Health Insurance to Achieve Universal Health Coverage. *BMJ.* 2019;365:2378. doi: 10.1136/bmj.l2378.

Julia Vedom et Huhua Cao, Health Care Access and Regional Disparities in China, *Espace populations sociétés*, 2011 January, 63-78.

Kim H., Mitra S. Dynamics of health and labor income in Korea, *The Journal of the Economics of Ageing*, Volume 21, 2022, 100351, ISSN 2212-828X, https://doi.org/10.1016/j.jeoa.2021.100351.

Meraya AM, Dwibedi N, Tan X, Innes K, Mitra S, Sambamoorthi U. The dynamic relationships between economic status and health measures among working-age adults in the United States. *Health Econ*. 2018;27(8):1160-1174. doi:10.1002/hec.3664

Nguyen L, Lee JT, Hulse ESG, Hoang MV, Kim GB, Le DB. Health Service Utilization and Out-of-Pocket Expenditure Associated with the Continuum of Disability in Vietnam. International Journal of Environmental Research & Public Health [Electronic Resource]. 2021;18(11):25.

National Bureau of Statistics. The Outcome of the Seventh National Population Census. 2021. Available online: http://www. stats.gov.cn/tjsj/sjjd/202105/t20210512_1817336.html (accessed on 6 December 2021).

Owens GM. Gender differences in health care expenditures, resource utilization, and quality of care. J Manag Care Pharm. 2008 Apr;14(3 Suppl):2-6. doi: 10.18553/jmcp.2008.14.S6-A.2. PMID: 18439060.

Prashant Loyalka & Lan Liu & Gong Chen &Xiaoying Zheng. Demography (2014) 51:97–118 DOI 10.1007/s13524-013-0272-7

Salinas-Rodríguez A, Manrique-Espinoza B, Mussot IT, Montañez-Hernández JC. Out-of-pocket healthcare expenditures in dependent older adults: Results from an economic evaluation study in Mexico. Frontiers in Public Health. 2020;8.

The State Council of the People's Republic of China (CPC) The Opinions of the Communist Party of China Central Committee and the State Council on Deepening Reform of the Health Care System. [(accessed on 17 December 2020)]; Available online: http://www.gov.cn/jrzg/2009-04/06/content_1278721.htm.

Wang N, Xu J, Ma M, et al. Targeting vulnerable groups of health poverty alleviation in rural China- what is the role of the New Rural Cooperative Medical Scheme for the middle age and elderly population?. *Int J Equity Health*. 2020;19(1):161. Published 2020 Sep 14. doi:10.1186/s12939-020-01236-x

Wooldridge, J. M. (2006). Introductory econometrics: A modern approach. Mason, OH: Thomson/South-Western.

World Health Organization. Ageing and health in China. WHO; 2019 Available online: https://www.who.int/china/healthtopics/ageing#:~:text=In%202019%2C%20around%2075%25%20of, cardiovascular%20disease%2C%20diabetes%20and%20hypertension

Xian W., Xu X., Li J., Sun J., Fu H., Wu S., Liu H. Health Care Inequality under Different Medical Insurance Schemes in a Socioeconomically Underdeveloped Region of China: A Propensity Score Matching Analysis. *BMC Public Health.* 2019;19:1–9. doi: 10.1186/s12889-019-7761-6.

You X, Kobayashi Y. Determinants of out-of-pocket health expenditure in China: analysis using China Health and Nutrition Survey data. Appl Health Econ Health Policy. 2011;9(1):39-49. doi: 10.2165/11530730-000000000-00000. PMID: 21174481.

Zhang A., Nikoloski Z., Mossialos E. Does Health Insurance Reduce out-of-pocket Expenditure? Heterogeneity among China's Middle-aged and Elderly. *Soc. Sci. Med.* 2017;190:11–19. doi: 10.1016/j.socscimed.2017.08.005.

Zheng PP, Guo ZL, Du XJ, Yang HM, Wang ZJ. Prevalence of Disability among the Chinese Older Population: A Systematic Review and Meta-Analysis. *Int J Environ Res Public Health*. 2022;19(3):1656. Published 2022 Jan 31. doi:10.3390/ijerph19031656

Zhao Y, Zhang P, Oldenburg B, Hall T, Lu S, Haregu TN, He L. The impact of mental and physical multimorbidity on healthcare utilization and health spending in China: A nationwide longitudinal population-based study. Int J Geriatr Psychiatry. 2021 Apr;36(4):500-510. doi: 10.1002/gps.5445

ESG Information and ESG Rating Updates

Kai Chen* and Dona Siregar[6]

ABSTRACT

This paper studies how firms' ESG information is incorporated into the stock market by examining ESG ratings and stock market returns. Using a buy-and-hold return analysis and a calendar-time portfolio analysis, we find that firms' ESG ratings are significantly associated with stock performance, suggesting that investors incorporate ESG information in the stock price accordingly. In an event study analysis, we further find evidence of significant market reactions measured by the cumulative abnormal returns around the days of ESG updates. The significance shows updates of ESG ratings convey some information that has not been perceived by investors and has yet to be integrated into the stock prices. Altogether, we provide additional evidence to support the importance of ESG ratings in the stock market.

Keywords: *ESG rating, ESG information, stock returns*

INTRODUCTION

ESG ratings are the principal source of investor information about firms' ESG behaviors. This study focuses on mechanisms of how firms' ESG information is incorporated in the stock market. Theoretically, there are two mechanisms for ESG information to be incorporated into stock prices. In one mechanism, investors are unwilling or unable to pay attention to firms' ongoing ESG practices, perhaps due to the prohibitive costs of ESG information gathering and processing. In this case, investors rely on firms' ESG information conveyed in ESG ratings and incorporate the information in firms' stock prices when the ESG ratings are published to the public. For these investors, the ESG ratings are indispensable and valuable. In the other mechanism, in which the stock market is highly efficient, investors keep track of firms' ESG practices over time and incorporate the ESG information in the stock prices accordingly. In this case, the periodically updated ESG ratings are less valuable to these investors because ESG information embedded in the ratings has already been incorporated into the stock prices prior to the release of the ESG ratings.

The results of this study suggest that investors incorporate ESG information into stock prices by combining these two mechanisms. Using Sustainalytics ESG ratings and applying the buy-and-hold return and calendar-time portfolio analysis, we find evidence suggesting that investors incorporate ESG information in the stock price as it becomes available. However, this information incorporation process may not be fully completed. Furthermore, we find that when ESG ratings are released, the updates in ESG ratings also convey some information that has yet to be perceived by investors, as evidenced by the

[6] Department of Business, State University of New York at Oneonta

significance of the cumulative abnormal returns around the ESG release days. In particular, the significant abnormal returns around the ESG rating updates are mainly driven by the information disclosure related to firms' social behaviors. We attribute the study's results as additional evidence to support the importance of ESG ratings in the stock market.

This study contributes to the literature that examines ESG information, ESG ratings, and stock returns. Studies on how stock markets react to positive and negative ESG events such Capelle-Blanchard and Petit (2019), Aoadi and Marsat (2018), and Krueger (2015) analyze a collection of ESG news (bad or good) or ESG controversies of publicly listed companies. In our study, we examine the regular updates of Sustainalytics ESG ratings published on Yahoo! and Bloomberg to analyze the relationship between ESG information and stock performance. Using the Sustainalytics ESG ratings allows our paper to provide additional insights into the role of ESG ratings on stock market returns to complement Glück et al. (2022) who examine market reactions of MSCI ESG rating change events.

PRIOR RESEARCH AND HYPOTHESES

Many investors rely on ESG rating information to make investment decisions, while corporations use ESG ratings to gain feedback on their sustainability initiatives. With sustainability's attention substantially increasing, so do investigations on the ESG ratings. ESG ratings bring transparency to companies about their sustainability efforts and condense it to a numerical score. Equally important to research on ESG ratings is whether the evidence shows ESG ratings are relevant to investors. Some evidence shows investors view ESG as value relevance. Serafeim and Yoon (2022a) examined how disagreements among three ratings (MSCI, Sustainalytics, and Thomson Reuters) predict future ESG news and the associated market reactions. They find that the consensus rating predicts future ESG news, and firms with considerable disagreements between raters have low predictive ability. ESG events such Capelle-Blanchard and Petit (2019), Aoadi and Marsat (2018), and Krueger (2015) analyze a collection of ESG news (bad or good) or ESG controversies of publicly listed companies. However, other studies find that ESG ratings do not correlate with stock performance and therefore do not provide useful information. Auer and Schuhmacher (2016) find that regardless of geographic region, industry, or ESG criterion, active selection of high- or low-rated stocks does not provide superior risk-adjusted performance compared to passive stock market investments. Hartzmark and Sussman (2019) do not find evidence high-sustainability mutual funds outperform low-sustainability funds. Existing reports and surveys (i.e., Boffo and Patalano, 2020; Amel-Zadeh & Serafeim, 2018) report barriers to ESG data integration into the investment decision processes.

In summary, some studies find evidence that ESG ratings are relevant for determining stock risks, returns, and future ESG rating predictability. In contrast, others find that using ESG ratings does not provide information. Studies on whether ESG ratings are valuable to investors have produced conflicting and incomplete results.

In this study, we examine the mechanisms of ESG information incorporation into stock prices. First, we examine if ESG information is vital to stockholders. We examine whether ESG behaviors relate to stock

performance. ESG rating positively related to the buy-and-hold abnormal returns provides evidence that ESG information is valuable to investors. Furthermore, given that ESG information is value relevant to investors, we propose two competing hypotheses.

The null hypothesis is that investors observe a firm's ESG practices daily routinely and integrate their practical information of the firm's ESG practices into their portfolio decision accordingly. If this hypothesis is confirmed, we expect CAR (cumulative abnormal return around the day of the ESG rating ranking publication) not to be related to the ESG ratings when the ESG ratings are updated. The competing hypothesis is that investors' valuation of a firm relies on the firm's ESG rating ranking. Investors may be unable or unwilling to observe the firm's ESG behavior themselves and use the ESG rating to make their portfolio decision. Nevertheless, they perceive that ESG rating is value relevant. Therefore, investors rebalance their portfolios accordingly whenever an ESG rating is published. If this hypothesis holds, we expect CAR (cumulative abnormal return around the day of the ESG rating publication) to be positively related to the ESG rating when the ESG rating is updated. The line between these two hypotheses might not be apparent in the real world. Nevertheless, it merits empirical analyses to test the hypotheses.

SAMPLE, DATA, and METHODOLOGY

Sample Selection

We start our sample with the 2966 companies in the Russell 3000 index as of January, 2014. Further selection requires companies to have ESG data and gives a result of 615 companies for the buy-and-hold abnormal return (BHAR) analysis and 578 firms for the cumulative abnormal return (CAR) analysis. Sustainalytics ESG rating is a rating to measure firms' sustainability for over 75,000 companies worldwide. Although Sustainalytics started releasing its ESG rating since 2014, the company began periodically updating the rating on Yahoo! Finance on the 5th day of every month since 2017. We use a sample of the firms in Russel 3000 with the Sustainalytics ESG rating available over the period from 2017 to 2018 to analyze the CAR of ESG release. We expand the sample period from 2014 to 2018 to include as much information as possible in the BHAR analysis. We only cover data until the end of 2018, and therefore do not include the new enhanced ESG Risk rating in the sample. Due to individual firm data availability in the Center for Research in Security Prices (CRSP) and Compustat databases, the number of companies in the sample reduces to 537 for the BHAR analysis and 496 for the CAR analysis. Following existing literature, we exclude depository institutions and utility companies from the sample as they are highly regulated industries. The final selection consists of 396 and 353 companies for the BHAR and the CAR analyses, consecutively.

ESG data

Sustainalytics is one of the leading providers of ESG ratings that use the framework of ESG as a measure how environmental, societal, and governance can financially material affect a company. Prior to 2019, Sustainalytics ESG rating evaluated a company concerning its general preparedness to deal with its

ESG risks and opportunities relative to other companies within the same industry. Bloomberg publishes the Sustainalytics ESG ratings each month that measures firms' corporate social responsibility (CSR) performance and practice of 9,000 to 11,000 publicly traded companies worldwide. The Bloomberg ESG ratings consist of three industry-percentile ranks: Sustainalytics Environments, Sustainalytics Social, and Sustainalytics Governance of each firm that Bloomberg publishes. In particular, the environmental score is determined by the level of environmental preparedness and disclosures, and environmental controversies. The social score is determined by the quality and controversies of policies, programs, and management systems regarding employees, suppliers, customers, and society. Finally, its governance score is determined by the practices related to board independence and elections, auditor independence, executive compensation, voting, and shareholder rights. The scores of environment, social, and governance are the bases for a percentile rank assigned to a company. The rank measures the company's environmental, social, and governance performance and practice compared to its industry peers. The ESG scores range from 0 to 100, representing the percentage of companies below a particular company's score. For example, a company's ESG score of 65.5 means 65.5% of firms in the same industry have lower ESG scores than the observed company. A similar interpretation applies to each environment, social, and governance score.

The Appendix shows the ESG variables used in the regression analysis. Summary statistics of the ESG variables show the average of RANK is 49.4 percentile, while the mean of GOV is 55 percentile. The average value of ENV is slightly lower at 46.9 percentile. The standard deviation of the ENV score is 29.2 percentile, while RANK and SOC are about 27 percentile. We use Sustainalytics ESG ratings available for the sample period from 2014 to 2018 to include as much information as possible for the BHAR analysis. The minimum value of difRANK is -40, meaning the largest one-month drop in RANK is 40 percentile. The maximum value of difRANK is 43.09, indicating the most one-month increase in RANK is 43.09 percentile. Looking at each component of the ESG rating, difGOV has the largest one-month drop and increase in the score (53.33 and 70.37, respectively).

ESG Behavior and Stock Price

We hypothesize that the information on a firm's ESG behavior is incorporated into the firm's stock market price. If ESG practices are relevant in the stock market, the information should be related to the stock price performance. We employ the Buy-and-Hold Abnormal Return (BHAR) to measure the stock price performance during a month. The BHAR is calculated using the following method:

$$BHAR = \prod_{t=1}^{T}(1 + R_{it}) - \prod_{t=1}^{T}\big(1 + E(R_{it})\big) \quad (1)$$

Where R_{it} is the actual return of the sample firm i on day t, and $E(R_{it})$ is the expected return of firm i on day t. $E(R_{it})$ is estimated with the market model following the standard setting of the WRDS. We calculate the BHAR of firm i over a calendar month from its first trading day to the last trading day of the month.

We analyze a firm's cumulative abnormal returns (CAR) to study market reaction to the release of ESG score. Cumulative abnormal return (CAR) over a 5-day window around 5th (15th) of t month, adjusted

by the market model. The 5th (15th) of the month is the day when Sustainalytics Company is supposed to update its monthly ESG ratings on Yahoo! Finance (Bloomberg). The actual update day might not be exactly this day.

The CAR over the five trading days around the release day is calculated using the following method:

$$CAR = \sum_{t=1}^{T}[R_{it} - E(R_{it})] \quad (2)$$

where R_{it} is the actual return of the sample firm i in day t, and $E(R_{it})$ is the benchmark return of firm i on day t. $E(R_{it})$ is estimated using the market model:

$$E(R_{it}) = \alpha_i + \beta_i R_{m,t} \quad (3)$$

where R_{mt} is the market portfolio return on day t and α_i and β_i are the firm i's alpha and beta that are estimated with the market model.

Correlation analysis shows the control variables have low correlations with other variables. The environmental and social rankings correlate more to the overall ESG RANK ranking than the governance ranking. No high correlations are found among the change in environment, social, and governance ranking.

RESULTS

BHAR and Calendar-Time Portfolio Analyses

In this section, we begin with testing the hypothesis that the information of firms' ESG practice has been incorporated into the stock market price before the monthly release of their ESG rankings. We test the hypothesis with multivariate regressions of the ESG ranking against the stock performance, controlling for other variables. We use the buy-and-hold abnormal (BHAR) return and the calendar-time portfolio methods for the firm stock performance analysis. The regressions are run with fixed effects model with AR(1) disturbance. We noted that ESG Sustainalytics ratings reflect the ESG practices of a firm during the month leading up to the release of the ESG ratings. The ESG ratings publicly released in month-t pertains to a firm's past month t-1 ESG practices. For this reason, we run multivariate regressions with the BHAR over t-1 month as the dependent variable, and the ESG ranking variables of month-t as the explanatory variables of interest.

Table 1 presents multivariate regression results on the one-month BHAR with the ESG ranking variables as the independent variables. We use the overall ESG ranking variable RANK as the independent variable in Model 1 and use the ESG characteristics separately in Models 2, 3, and 4 because they are highly correlated, as reported in the correlation matrix table (Table 3). The results show that BHARs positively associate with firms' ESG rankings. Firms with higher rank in ESG have a significantly higher buy-and-hold abnormal return (Model 1). In addition, all three ESG components that make up the overall ranking have a significant association with the monthly buy-and-hold abnormal returns when analyzing them separately. The results suggest that ESG ranking that is released in month-t is highly related to the stock market performance in month t-1. The stock price return of month t-1 contains the ESG behaviors of month-t.

Table 1: Analysis of BHAR over the 2014 – 2018 Sample

	(1)	(2)	(3)	(4)
VARIABLES	**BHAR**	**BHAR**	**BHAR**	**BHAR**
RANK	0.0003***			
	(3.97)			
ENV		0.0002***		
		(3.50)		
SOC			0.0002***	
			(3.14)	
GOV				0.0001**
				(2.33)
2017	0.0017	0.0015	0.0019	0.0021
Control variables	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes
Obs.	19,434	19,434	19,434	19,434
# of Firms	396	396	396	396

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

We further explore the calendar-time portfolio analysis to evaluate the long-term stock performance in relation to firms' ESG. This method controls possible cross-sectional correlation among firms when examining long-term performance to address overstated t-statistics caused by the independency of firms' abnormal returns over time (Fama 1998; Mitchell and Stafford 2000). An observation in the sample is marked as 1 if its RANK, ENV, SOC, or GOV is above the 50th percentile and otherwise as 0. For each ESG variable (RANK, ENV, SOC, or GOV), two portfolios - Above the 50th Percentile and Below the 50th Percentile - are constructed. Each month, a stock is included in one of the two portfolios according to its current month's ESG rating. Then the portfolios' return of each month is calculated based on equal-weighted portfolio return (EW), which is the arithmetic mean of stock returns in the portfolio in each month. We calculate the monthly excess return of this moving portfolio by subtracting the risk-free return from the portfolio return. The time-series portfolio excess returns are regressed on the three Fama and French (1993) factors plus the momentum factor proposed by Carhart (1997). The equal-weighted portfolio returns minus risk-free rate is the dependent variable. To capture the difference in the abnormal return between the two portfolios, we add a dummy variable, PTF.

$$RP_t - RF_t = \alpha + \beta_1(RM_t - RF_t) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 UMD_t + \gamma PTF + e_t \quad (3)$$

In this model, $RP_t - RF_t$ is the return of the portfolio in excess of the one-month T-bill rate. The four independent variables are the excess return of the market portfolio, ($RM_t - RF_t$); the difference in returns between a small stock portfolio and a big stock portfolio, SMB; the difference in returns between a high book-to-market stock portfolio and a low book-to-market stock portfolio, HML; and the difference in returns between a high prior momentum stock portfolio and a low prior momentum stock portfolio, UMD. The intercept, α, measures the average monthly abnormal returns on the portfolio of Below the 50th Percentile. The coefficient, γ represents the returns on the portfolio of Above the 50th Percentile in excess of the return on Below the 50th Percentile.

Table 2: Calendar-Time Portfolio Analysis of Stock Performance over the 2014-2018 Sample

Using equal-weighted returns of the portfolio as the dependent variable			
	(1)	**(2)**	**(3)**
	Above 50th	**Below 50th**	**Difference**
RANK	0.0013	-0.0011	0.0027**
	(1.51)	(-1.10)	(2.13)
ENV	0.0013	-0.0098	0.0026*
	(1.47)	(-1.00)	(1.91)
SOC	0.0012	-0.0009	0.0022*
	(1.48)	(-0.92)	(1.85)
GOV	0.0009	-0.0011	0.0022*
	(0.94)	(-1.15)	(1.70)

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

In Table 2, Columns (1) and (2) present the intercepts of α, estimated with the model (3) without the dummy variable of *PTF* on the two sub-samples, the Above 50th Percentile and the Below 50th Percentile, respectively. The result that the intercepts, measuring the average monthly abnormal returns, are not significantly different from zero support the efficient market hypothesis. Column (3) is the estimated coefficient γ of the dummy variable PTF, which equals 1 if an observation has RANK, ENV, SOC, or GOV above the 50th Percentile and 0 otherwise when the four-factor regression is applied to the entire sample. The coefficient of the dummy measures the difference in stock performance between the two portfolios along with the t-statistics. The results show that firms with higher ESG ratings have higher average abnormal returns with moderate statistical significance at the 5% and 10 % confidence levels.

We conclude that results of the BHAR and calendar-time analyses support the case that the stock price incorporates ESG information, at least partially significant. We further examine whether the information has been fully included in the stock price performance.

Event Study

Table 3 presents the results of the 5-day CAR analysis around the fifth day of each month for the 2017-2018 sample. The multivariate regressions estimate the relationship between the cumulative abnormal returns over five days around the 5th day of the month (CAR5th) and the ESG ratings. The

regressions are run with fixed effects model with AR(1) disturbance. We used samples during 2017-2018 when Sustainalytics was supposed to regularly update its ESG Ratings on Yahoo! Finance on the 5th of the month. We use variables that measure the change of ESG ranking of month-t from the month t-1 ESG ranking (" dif") as the independent variables. The ESG change variables may contain information investors have not yet incorporated into the stock price. They measure unexpected changes in the ESG scores.

Results show that none of the changes in the ESG score rankings is significant. They suggest the market has anticipated the changes and reflected them in the month t-1 stock performance. The 5th-day CAR analysis results support the hypothesis that the market is efficient. However, these results may need to be revisited for two reasons. (1) the ESG ranking data are collected from Bloomberg, which released them on the 15th day of the month. The ranking is built from Sustainalytics raw ESG score data that are released on the 5th day of the month on Yahoo! Finance. The ranking orders might be different from the raw ESG scores. Thus, the 5th CAR results might be insignificant because of the differences. (2) Investors might not pay attention to the release of raw ESG scores on Yahoo! Finance, so the 5th CAR results are insignificant. If this is the case, it suggests investors keep track of firms' ESG behavior in month t-1, and do not keep track of the release of the ESG rating. However, this notion that investors track all ESG information may be disputable due to the high costs associated with gathering every ESG news without the help of intermediaries such as Sustainalytics for gathering ESG information.

Table 3: Analysis of 5-Day CAR around 5th Day over 2017-2018 Sample

	(1)	**(2)**	**(3)**	**(4)**
VARIABLES	**CAR5th**	**CAR5th**	**CAR5th**	**CAR5th**
difRANK	0.0002			
	(1.53)			
difENV		-0.0001		
		(-1.02)		
difSOC			0.0002	
			(1.47)	
difGOV				0.0001
				(1.03)
RANK(-1)	0.0001			
	(1.63)			
ENV(-1)		-0.0000		
		(-0.38)		
SOC(-1)			0.0001**	
			(2.26)	

GOV(-1)				0.0001
				(1.01)
Control variables	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes
Obs.	7,409	7,409	7,409	7,409
# of Firms	353	353	353	353

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

The above reasonings justify us moving forward to the examination of the 15th day of the month for Bloomberg to release the ESG rating. We further do a robustness check by using the 15th day of the month for the CAR analysis.

Table 4 shows that the environmental ranking (difENV) is insignificantly related to the five days of cumulative abnormal returns on the 5th day of the month. Therefore, when the ESG score ranking is released, the change in the environmental ranking from the previous month does not correlate with abnormal returns. The insignificance of the coefficient implies that the release of the environmental score ranking does not contain new information. Similarly, it suggests that the environmental information of month t-1 has been incorporated into the previous month's stock prices.

Turning to difSOC, we observe that the difSOC coefficient is significant. When investors observe the social practices of firms in the past month, they may only perceive partial information about the firms' social practices. Consequently, the stock prices only incorporate such partial information. However, when Sustainalytics releases a new SOC score in the following month, the release of the score discloses information that has yet to be incorporated into stock prices; the information is new to the investors. As a result, the disclosure of the SOC score generates abnormal returns over the event window. For example, one unit increase in the percentile of the SOC score would increase the cumulative abnormal returns by 0.02% over the five days event window.

Table 4: Analysis of 5-Day CAR around 15th Day over the 2017-2018 Sample

	(1)	**(2)**	**(3)**	**(4)**
VARIABLES	**CAR15th**	**CAR15th**	**CAR15th**	**CAR15th**
difRANK	0.0003**			
	(2.40)			
difENV		0.0001		
		(0.92)		
difSOC			0.0002**	
			(2.30)	

difGOV				0.0001
				(1.31)
RANK(-1)	0.0002***			
	(3.08)			
ENV(-1)		0.0001**		
		(2.06)		
SOC(-1)			0.0002***	
			(3.45)	
GOV(-1)				0.0001
				(1.18)
Control variables	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Intercept	Yes	Yes	Yes	Yes
Obs.	7,409	7,409	7,409	7,409
# of Firms	353	353	353	353

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

The CAR analysis also shows that difRANK is significantly correlated to abnormal returns. The RANK score is built to combine the three components of ESG into one metric that comprehensively measures firms' ESG practices. Along with the significance of difSOC, the significant coefficient of difRANK is driven by the release of the SOC score. An increase in one unit percentile of RANK would contribute to 0.03% abnormal returns during the event window analysis.

CONCLUSION

In this paper, we examine how Bloomberg ESG ratings relate to stock market returns to study how firms' ESG information is incorporated into their stock market performance. Our results show that market participants capture firm ESG information timely and adjust their trading behavior accordingly. Nevertheless, this information incorporation is incomplete. On average, updates of firms' ESG ratings still contain some information about ESG practices the market has not perceived. Overall results of the study indicate that the ESG ratings are supplementary but indispensable in facilitating information to the stock market.

ENDNOTES

1. We are grateful for helpful suggestions from Joan Nix, Elerna Smirnova, and participants at the 2022 annual conference of the New York State Economics Association, Old Westbury, NY.

2. Complete summary statistics, correlations, and regression results are available from the authors upon request.

REFERENCES

Amel-Zadeh, A. and Serafeim, G. 2018. Why and how investors use ESG information: Evidence from a global survey. *Financial Analysts Journal*, 74(3), pp.87-103.

Aoadi, A. and Marsat, S. 2018. Do ESG controversies matter for firm value? Evidence from international data. *Journal of Business Ethics*, 151(4), pp.1027-1047.

Auer, B.R. and Schuhmacher, F. 2016. Do socially (ir) responsible investments pay? New evidence from international ESG data. *The Quarterly Review of Economics and Finance*, 59, pp.51-62.

Boffo, R. and Patalano, R., 2020. ESG investing: Practices, progress and challenges. *Éditions OECD, Paris.*

Capelle-Blancard, G. and Petit, A. 2019. Every little helps? ESG news and stock market reaction. *Journal of Business Ethics*, 157, pp.543-565.

Carhart, M.M., 1997. On persistence in mutual fund performance. *Journal of Finance*, 52: 57– 82.

Fama, E.F. 1998. Market efficiency, long-term returns, and behavioral finance1. *Journal of Financial Economics*, 49(3): 283-306.

Fama, E. F., & French, K. R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1): 3-56.

Glück, M., Hübel, B. and Scholz, H. 2021. ESG rating events and stock market reactions. *Available at SSRN 3803254*.

Hartzmark, S.M. and Sussman, A.B. 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6), pp.2789-2837.

Krüger, P. 2015. Corporate goodness and shareholder wealth. *Journal of Financial Economics*, 115(2), pp.304-329.

Mitchell, M. L., & Stafford, E. 2000. Managerial decisions and long-term stock price performance. *Journal of Business*, 73(3): 287-329.

Serafeim, G. and Yoon, A. 2022a. Stock price reactions to ESG news: The role of ESG ratings and disagreement. *Review of Accounting Studies*, pp.1-31.

Serafeim, G. and Yoon, A. 2022b. Which corporate ESG news does the market react to?. *Financial Analysts Journal*, 78(1), pp.59-78.

Appendix: Variable Definition

Variable	Definition
RANK	A comprehensive rating that measures a firm's environment, social, and governance practice and status over t-1 month, shown in percentile in industry, released on Bloomberg around 15th day of t month
ENV	A rating that measures a firm's environment practice and status over t-1 month, shown in percentile in industry, released on Bloomberg around 15th day of t month.
SOC	A rating that measures a firm's social practice and status over t-1 month, shown in percentile in industry, released on Bloomberg around 15th day of t month.
GOV	A rating that measures a firm's governance practice and status over t-1 month, shown in percentile in industry, released on Bloomberg around 15th day of t month.
The change in ESG rankings	difRANK, difENV,difSOC,or difGOV. A variable that measure ESG rank of month-t minus the month t-1 ESG rank.
2017	Dummy variable: one if an observation occurs after January 1st, 2017, zero otherwise. Since 2017, the monthly E/S/G ratinges are posted on Yahoo Finance around the 5th of the following month.

Notes: (1) A variable followed by (-1) indicates that the data used for this variable are from the previous month. For example, Beta(-1) represents the previous month's Beta. (2) All the data used to calculate the Compustat variables come from the latest annual report prior to the year of t month.