

RISK AND REWARD: STORIES FROM BOTH CATASTROPHE BOND MARKET AND
VENTURE CAPITAL INVESTMENT STRATEGIES

By

Yutian Li

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Business Administration—Finance—Doctor of Philosophy

2022

ABSTRACT

RISK AND REWARD: STORIES FROM BOTH CATASTROPHE BOND MARKET AND VENTURE CAPITAL INVESTMENT STRATEGIES

By

Yutian Li

The first essay of this dissertation is about the Catastrophe bond. The catastrophe bond market is a good place to reveal how investors think about the climate risk in the mid-term foreseeable future. By studying catastrophe bond market transaction data, I find out that investors treat climate related bonds and non-climate related bonds differently. The yield difference between short-term climate related bonds and mid-term climate related bonds is significantly higher than the yield difference between short-term non-climate related bonds and mid-term non-related bonds. This yield difference is not caused by annual climate seasonality, different local property value growths, and different term structures. Tests on the implied hazard rate of catastrophe bonds show similar results. I also find that the difference in yield difference between climate related bonds and non-climate related bonds gets larger after some major climate related disaster events happened between the year 2017 and the year 2018, which indicates that this difference could be connected to the investors' perception of the uncertainty of future climate situations.

The second essay of the dissertation is related to venture capital. It addresses the necessity and benefits of expanding venture capital fund's investment selection pool geographically. We find out that funds with high expertise concentration levels tend to invest further away than funds with relatively low expertise concentration levels, especially when funds are not located in California. What's more, our results show that funds with larger geographic coverage outperform others, and funds with high expertise concentration levels outperform funds with relatively low expertise concentration levels in general. The out-performance is consistent across all funds'

spectrum. Last but not least, we find that funds' faraway investments outperform their own nearby investments in terms of excess IPO rate and excess fail rate, which can provide a useful guideline for venture capital funds.

This dissertation is dedicated to my parents, Changqing Li and Bilan Hu.

ACKNOWLEDGEMENTS

I thank my dissertation committee – Mark Schroder, Zsuzsanna Fluck, Ryan Israelsen, and Hao Jiang - for their guidance throughout my time at Michigan State. I owe special thanks to both Mark Schroder and Zsuzsanna Fluck, the chairs of my committee, for incredible mentorship.

I also thank the rest of the finance department faculty, especially Morad Zekhnini, for helpful comments and job market advice.

Finally, I thank my family and friends for endless love and support.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	x
CHAPTER 1. Mid-term Climate Change Risk in Catastrophe Bond Market	1
1. INTRODUCTION	2
2. CATASTROPHE BOND.....	7
3. DATA DESCRIPTION	9
3.1. Data Sources and Sample Selection.....	9
3.2. Variables	10
3.3. Summary Statistics.....	13
4. MID-TERM CLIMATE RISK PREMIUM.....	16
4.1. First Identification.....	16
4.2. Potential Problems	21
4.3. Robustness Test on Seasonality	22
4.4. Robustness Test on Term Structure	27
4.5. Robustness Test on Housing Price.....	31
4.6. Implied Hazard Rate	33
5. AN EVENT STUDY	38
6. DISCUSSION AND CONCLUSION.....	42
APPENDICES	43
Appendix A: Seasonality Adjusted Annual Expected Loss Rate	44
Appendix B: Average Implied Hazard Rate	48
BIBLIOGRAPHY.....	49
CHAPTER 2. Long Distance? Not That Bad: Venture Capitalists' Geographic Coverage and Their Investments.....	52
1. INTRODUCTION	53
2. DATA DESCRIPTION	57
2.1.Data.....	57
2.2. Concentration Index.....	57
2.3. Control Variables	59
2.4. Dependent Variables and Benchmark Exit Rate.....	61
2.5. Summary Statistics.....	63
3. FUND'S GEOGRAPHIC COVERAGE	64
3.1. Concentration Index and Geographic Expansion	64
3.2. California Funds and California Investments	65
3.3. Identification	69
4. GEOGRAPHIC COVERAGE AND PERFORMANCE.....	74
4.1. Identification	74
4.2. Fund's Faraway and Nearby Investments	81
5. DISCUSSION AND CONCLUSION.....	84

APPENDICES	86
Appendix A: Soft Cosine Similarity Score	87
Appendix B: Exit rates states distribution	89
Appendix C: Exit rates states distribution	90
Appendix D: Fund's Investment Geographic Coverage and Fund's Exit Performance	92
BIBLIOGRAPHY	94

LIST OF TABLES

Table 1: Basic Statistics.....	14
Table 2: First Identification.....	19
Table 3: Example for Seasonality.....	21
Table 4: Robustness Tests on Seasonality Part A.....	24
Table 5: Robustness Tests on Seasonality Part B.....	26
Table 6: Robustness Tests on Term Structure Part A.....	29
Table 7: Robustness Tests on Term Structure Part B.....	30
Table 8: Robustness Tests on Housing Price.....	32
Table 9: Implied Hazard Rate Part A.....	36
Table 10: Implied Hazard Rate Part B.....	37
Table 11:Recent Disaster Events.....	39
Table 12: Tests For The Event Study.....	41
Table 13: Exit Distribution.....	62
Table 14: Fund Type and Its Investment Distribution.....	63
Table 15: Concentration Index on Geographic (Domestic Investments).....	71
Table 16: Concentration Index on Geographic (Overall Distance).....	73
Table 17: Geographic Coverage and Excess Rates.....	77
Table 18: Exit Performance: California VS non-California.....	78
Table 19: Exit Performance: Size, Experience, and Expertise.....	78
Table 20: Exit Performance (Among Funds): Faraway VS Nearby.....	80
Table 21: Exit Performance (Inside fund): Faraway VS Nearby.....	84

Table 22: Company Exit rate by states.	89
Table 23: Faraway Investment Ratio And Concentration Index	90
Table 24: Nearby Investment Ratio And Concentration Index	91
Table 25: Excess Exit Rates And Faraway Investment Ratio.....	92
Table 26: Excess Exit Rates And Nearby Investment Ratio.....	93

LIST OF FIGURES

Figure 1: Climate Related Bonds.....	15
Figure 2: Non-climate Related Bonds.....	15
Figure 3: Landfalling.	21
Figure 4: Fund Investment Year Span Distribution.....	60
Figure 5: Percentage of All Portfolio Companies.....	66
Figure 6: Investments Distribution by States.....	68

CHAPTER 1. Mid-term Climate Change Risk in Catastrophe Bond Market

1. INTRODUCTION

"Climate change is not priced into markets but its effect could be substantial, experts say", a headline from a CNBC news report on February 24th, 2021. In the report, Marchel Alexandrovich says " short of a climate disaster, (climate change) is a problem, but a slow storm that's brewing." But is that right? They may need to look into the catastrophe bond market to find out.

Climate change risk is seen as a long-term risk in the eyes of many investors. The potential economic damage from climate change as a long-term risk is hard to estimate because there are a lot of time-varying factors that can affect the climate condition in the long-term. What's more, the direct economic impact of those potential long-term risks, such as rising sea levels and higher than average temperatures, are not easy to be seen on the foreseeable companies' income statements. Therefore, investors may not pay enough attention to the danger of the climate risk, and it is not easy to identify the climate change risk from some mainstream financial markets. But climate related disasters, such as hurricanes, floods, and wildfires, are happening every year, and the economic damage from them can be known directly in a relatively short time. The climatic risk from these natural disasters is not only long-term but is also much closer mid-term. The main question is whether those mid-term climate change risks are recognized by the investors.

The catastrophe bond market faces these natural disasters directly, such that their transaction data can effectively reflect how investors really think about the risk of climate change. Details on catastrophe bonds are shown in section 2. One drawback about catastrophe bonds is that the typical years to maturity of a newly issued catastrophe bond is 3 to 4 years. Therefore, the catastrophe bonds are only able to capture a near future mid-term climatic disaster risk. One may say that the climate will not change on a noticeable scale in 5 years. While this might be true for

average sea levels or the average rise in temperature, for climate related disasters like hurricanes, floods, and wildfires, things could escalate very quickly.

There is a lot of evidence showing that the intensity of these disasters is increasing over time. The major tropical cyclone exceedance probability increases by 8% per decade, with a 95% CI of 2 to 15% per decade (Kossin, Knapp, Olander, and Velden, 2020). The proportion of category 4 and 5 hurricanes has increased at a rate of 25 to 30 % per degree of global warming (Holland, Bruyere 2013). Three of the top four costly hurricanes in the history of the United States happened in 2017 from August to September alone. Human-caused climate change contributed to an additional 4.2 million hectares of forest fire area during 1984–2015, nearly doubling the forest fire area expected in its absence (Abatzoglou, Williams 2016). The Australia wildfire in 2019 and the wildfires in the United States in 2020 showed us the scale of these disasters. One extra category 5 hurricane would cause unimaginable damage in a very short time once it hits landfall in a city area. And the probability of having an extra category 5 hurricane is not that small in a 5-year period. Therefore, climate change risk is not that far away from us, and it is important to understand how investors value this mid-term risk.

There are a lot of papers studying catastrophe bond markets but most of them are focusing on the pricing of the bond from the issuer's perspective like Ma and Ma (2012), Galeotti, Gurtler, and Winkelvos (2013), and Cox and Pedersen (2000). Those pricing papers have not addressed the difference between different types of natural risk. In the catastrophe market, about 1/3 of the bonds are only covering climate related risk, and about 1/4 of the bonds are only covering non-climate related risk. This paper will address the risk difference between climate related bonds and non-climate related bonds. The risk difference between those bonds can reveal investors' ideas on the uncertainty of near-future climate change risks.

The basic idea is that investors will price the future climate change risk into the catastrophe bonds that cover climate related disasters. For catastrophe bonds, I first put hurricanes, wildfires, and floods as well as related disasters into a climate related category and put others into a non-climate related category. There are two main differences between climate related bonds and non-climate related bonds. The first difference is that climate related bonds in this paper will experience seasonality every year, just like hurricane and wildfire seasons. The second difference is that climate related bonds are facing potential climate change risk, which is the main focus of this paper. In this paper, I assume climate will not change in the short term (at most less than one and half years). In order to reveal the climate change concern for the investors, I divide all the bond transactions data into two groups, short-term and mid-term, based on the predictability of their covered disasters and years to maturity. Therefore, I have 4 subgroups of observations: climate related short-term bonds, climate related mid-term bonds, non-climate related short-term bonds, and non-climate related mid-term bonds.

By conducting difference-in-difference tests on these four subgroups of observations, I find out that investors will ask for a higher yield for longer term climate related bonds when compared with other groups of catastrophe bonds. The yield difference between short-term climate related bonds and mid-term climate related bonds is about 150 to 200 basis points more than the yield difference between short-term non-climate related bonds and mid-term non-related bonds. One more year increase in years to maturity will increase the difference by 50 to 78 basis points. Considering 90% of the excess yields are between 0.0195% to 0.115%, the economic magnitude is relatively large. With the help of seasonality adjusted annual expected loss rate, I am able to decrease the effect from climate seasonality and provide more evidence that supports the results from previous tests. By assuming the coefficient on expected loss is a time-increasing variable, I

effectively rule out the potential term structure difference between climate related bonds and non-climate related bonds and provide more evidence to support the existence of the mid-term climate risk premium. I add the state level Zillow Home Value Index annual growth rate into the regression to control for local real estate market growth, and I still get similar results as the first test.

In order to have a more direct view of how investors sense the near future climate risk, I use the implied hazard rate to conduct the same difference-in-difference tests as I have done on excess yield to maturity. The implied hazard rate difference between short-term climate related bonds and mid-term related bonds is about 260 to 340 basis points more than the implied hazard rate difference between short-term non-climate related bonds and mid-term non-related bonds. One more year increase in years to maturity will increase the implied hazard rate difference by 90 to 120 basis points. Again, the magnitude of this difference is relatively large. Even after seasonality adjustment and controlling the local property value growth, the implied hazard rate still shows similar results as they were before.

So far, all the tests point out that there is a yield difference between climate related bonds and non-climate related bonds, and all the tests indirectly relate the difference to near-future climate change risk. There are two channels that can explain this difference. One is the mid-term prediction of climate change, and the other is the investor's increased risk aversion towards the uncertainty of climate change risk. The first channel is hard to test due to the lack of data and the unpredictability of the longer term climate situation. I use an event study and find some evidence to support the second channel. From 2017 to 2018, there were 7 major climate related disasters, which cost more than 700 billion dollars and caused 22 climate related bonds to default. If the second channel can contribute to the mid-term climate risk premium, then we can imagine that this mid-term climate risk premium will get even larger after those extreme disasters happened. I use

the transaction data from the bonds that were not affected by these 7 events to test how investors react to those extreme climate disasters. By using difference-in-difference-in difference tests, I find out that during those climate event periods, the yield difference increased by 75 basis points when compared with the yield difference outside the disaster period. One more year increase in years to maturity will increase the difference by 46 basis points. The yield difference increased on the bonds unrelated to those disaster events which provides some evidence that investors would increase their degree of risk aversion to future climate risk after some major climate related disaster happened. Therefore, the mid-term climate risk premium from previous tests could come from investors' increased risk aversion toward the future uncertainty of climate risk.

This paper is related to the climate change risk literature. Many studies in the literature paid attention to how fund managers react to these climatic disasters. For example, because of the salience bias, managers within a major disaster region underweight disaster zone stocks more than distant managers (Alok, Kumar, Wermers 2020). But they didn't identify investors' perceptions of future climate risk. Some research tried to identify the future climate risk, but whether the financial market will price it effectively somewhat depends on investors' beliefs and perceptions of climate change. From Yale climate opinion maps, most people think global warming is happening and climate change risk is real. In the real estate market, some evidence points out that houses projected to be underwater in climate change believer neighborhoods sell at a discount compared to houses in climate change denier neighborhoods (Baldauf, Garlappi, and Yannelis 2019). There is a systematic variation in how institutional investors see the importance of climate change risk based on their characteristics and beliefs (Ilhan, Krueger, Sautner, and Starks 2020). In the stock market, some retail investors revise their beliefs about climate change upward and sell carbon-intensive firms when experiencing abnormal warm temperatures in their area (Choi, Gao, and Jiang 2019).

So far most of the current research is focused on the long term risk premium from climate change risk such as sea levels rising and higher than average temperatures. But investors cannot experience direct damage from those potential long term risks. And peoples' opinions on long term climate change risk can change over time since the long term risk depends on how we as a society deal with the environment. Those are possible reasons why in some research there is no climate change effect detected in some markets. For example, research on establishment data from 1990 to 2015 found little evidence that temperature exposures significantly affect establishment-level sales or productivity, including among industries traditionally classified as 'heat sensitive' (Addoum, Ng, and Bobea 2019). Further, Murtin and Spiegel (2019) find limited housing price effects from future climate risk by using different datasets and methods.

This paper finds out that sophisticated institutional investors do pay an attention to the mid-term climate change risk in the catastrophe bond market and connects the mid-term climate risk premium to the unpredictability of near-future climate related natural disaster events. Moreover, this work also can help with the catastrophe bond price theory by introducing the climate change risk.

2. CATASTROPHE BOND

Catastrophe bonds are risk-linked securities that transfer a specified set of natural disaster risks from a sponsor to investors. Most catastrophe bonds are quarterly floating rate coupon bonds with different benchmark risk free rates like libor 3m, libor 6m, t-bill 1m, t-bill 3m, etc. Typical years to maturity for a newly issued catastrophe bond is about 3 to 4 years. All catastrophe bonds are issued at par value except zero coupon catastrophe bonds. Most investors are institutional investors including catastrophe funds, mutual funds, hedge funds, and reinsurers.

The proceeds raised from investors go to a special purpose vehicle which will invest those proceeds into a low-risk market. SPV will pay investors coupons periodically and principal at the end of the maturity of the bond if there is no trigger event happening during the bond's lifetime. A catastrophe bond will experience principal loss once it gets triggered by its covered natural disaster. The size of the loss on principle depends on the economic (or physical) magnitude of the disaster. Each catastrophe bond has a specific trigger mechanism.

There are five major trigger mechanisms.

1. Indemnity: the bond will be triggered if the issuer's actual losses from its covered disasters surpass a certain level.
2. Modeled loss: the bond will be triggered if the issuer's modeled loss from its covered disasters surpasses a certain level. Normally it is going to take a long time for the issuer to determine the real loss from the disaster. In order to shorten the time, the issuer use parameters from the disaster to simulate the loss (modeled loss) from the disaster.
3. Indexed to industry loss: the bond will be triggered when the insurance industry loss surpasses a certain level because of this disaster event.
4. Parametric: the bond will be triggered if the disaster's parameters (windspeed, ground acceleration, etc.) surpass a certain level.
5. Parametric index: a combination of modeled loss and parametric.

It is important to know that cat bonds that don't use a parametric trigger normally will reset its trigger line annually to maintain the same risk exposure level to its investors.

In our sample of tests, 73.1% of the bonds use the Indemnity trigger mechanism, 11.9% of the bonds use the Indexed to industry loss trigger mechanism, 6.2% of the bonds use the parametric trigger mechanism, and 8.8% of the bonds use Modeled loss trigger and other trigger mechanisms.

Let's take Citrus Re 2015-1 B as a sample. It was issued on April 8th, 2015 and is expected to mature on April 9th, 2020. The size of Citrus Re 2015-1 B was 97.5 million US dollars. It covers U.S. named storms in Florida. It was a quarterly floating rate coupon bond initially with a coupon formula as $6\% + 3\text{-month t-bill rate}$. It used indemnity as the trigger mechanism. The bond was triggered on May 17th, 2018 for the first time due to Hurricane Irma in 2017. The principal payment dropped 5 times from 100% to 50.5% gradually from May 17th, 2018 to July 17th, 2019 as the loss from Hurricane Irma became clear. The fixed part of the coupon dropped from 6% to 0.5% on April 9th, 2018 till the end of the maturity. The bond changed to a monthly floating coupon bond on April 9th, 2018 and used 1 monthly t-bill rate as its floating part of the coupon. Once a Catastrophe bond is triggered by its covered disaster, the coupon structure and the principle normally will change dramatically. But in another case, like Citrus Re 2016-1 D, the fixed part of the coupon rate didn't change even though its principle dropped to 64.7%.

3. DATA DESCRIPTION

3.1. Data Sources and Sample Selection

There were around 560 catastrophe bonds issued between January 2012 to March 2022 globally. I collect offering deals from www.artemis.bm. The website offers annual expected loss and annual attachment rate for most of the bonds. More bond-specific information, including bond offering date, expected maturity date, coupon structure, size of the bond, benchmark risk free rate, and default information, are from Bloomberg. All the bond transaction price data are from TRACE.

Among those 560 catastrophe bonds, 356 have complete information including both transaction price and bond specific characteristic data. Out of those 356 bonds, 193 of them are single disaster risk covered bonds. 113 of those 193 bonds are climate related and the remaining

80 bonds are non-climate related. A bond is defined as being climate related if its underlying natural disaster risk is related to hurricanes, typhoons, wind storms, floods, or wildfires. A bond is non-climate related if its underlying natural disaster risk is due to earthquakes or extreme morbidity (Medical benefit claims levels). Most of the bonds cover natural disasters in the US, and some of the bonds cover natural disasters that happen in Australia, Canada, EU, Japan, Mexico, and Turkey. All the bonds in my sample are floating rate coupon bonds except five zero coupon bonds. 134 of all floating rate coupon bonds are quarterly and the remaining 54 are monthly. The coupon bonds here have a fixed part and a floating part. The floating part of the coupon is a benchmark risk free rate which can be a three-month libor rate, six-month libor rate, three-month t-bill rate, or others. Time-varying benchmark risk free rate (floating part of the coupon) for each bond from coupon date A to coupon date B is determined at coupon date A.

3.2. Variables

A majority of the tests use excess yield and implied hazard rate as the dependent variable. Excess yield is defined as a bond's yield to maturity minus the bond's specific risk free rate. Excess yield of bond i at time t is $y_{it} = ytm_{it} - r_{it}$. Compared with yield to maturity, excess yield is more accurate to reflect the risk from the natural disaster because risk free rate changes over time. Implied hazard rates are backed out from the bond's transaction prices which is explained in the appendix. Details on how to back out the implied hazard rates are given in the appendix. Yield to maturity is calculated from

$$price = \sum_{a=1}^N \left(\frac{(coupon + r_a) / k}{(1 + ytm)^{t_a}} \right) + \frac{FV}{(1 + ytm)^{t_N}}$$

price is the transaction price, and coupon is the annual fixed part of the coupon rate for the bond. k is the frequency of the coupon, N is the total number of coupons remaining for the bond after the transaction, FV is the face value of the bond, t_a is the time period between the time investors will receive their a th cash flow and the time they buy the bond, and r_a is bond's benchmark risk free rate at bond's a th payment date. When $a = 1$, r_a is the spot rate of the bond's specific benchmark risk free rate at the bond's last payment date, and when $a > 1$, r_a is the forward rate of the bond's specific benchmark risk free rate which is determined at the end of the transaction date. The difference between t_{a+1} and t_a are not necessarily the same because of the transaction time and the irregularity of the coupon payment schedule. All the spot rates and forward rates are also from Bloomberg. In order to distinguish between short-term and mid-term bonds, I define mid-term bonds as the bond that will still experience at least two hurricane seasons. In this definition, for mid-term climate related bonds the investor cannot effectively predict the climate situation for the next and future hurricane seasons. This uncertainty on the future climate situation will be the main subject tested in this paper. Both mid-term dummy variable and years to maturity will be used as the maturity measurement, which is denoted as M , in the tests. Unlike the non-climate related bonds, climate related bonds have seasonal risk exposure. All the climate related bonds have almost the same climate season every year in the specific region. Bonds that cover the northern hemisphere, will experience a climate season from June to November. On the contrary, bonds that cover the southern hemisphere, like Australia, will experience a climate season from November to April next year. During the non-climate season period, the risk of experiencing a major disaster is extremely low and will be considered zero in this paper. To control the potential seasonal effect on investors' perception on climate related risk, I use the climate season dummy variable which equals one if the transaction of a bond happens during a climate season period and is zero

otherwise.

The risk level of each bond in this paper is represented by the annual expected loss rate, Conditional expected recovery rate, and coupon rate. The annual expected loss rate, denoted EL is offered by the catastrophe bond pricing companies along with the attachment rate, denoted as AR . That information can be found on the website www.artemis.bm. The expected loss in this paper is the annual expected percentage loss on par value. The attachment rate of a catastrophe bond is the probability of that bond getting triggered by its covered disaster in one year. Need to notice that, for both parametric trigger and indemnity trigger related bonds, EL and AR will be the same value over the years because bond issuer normally will reset their event trigger line annually. Conditional expected recovery rate, denoted as R , can be deduced from $EL=AR \times (1-R)$. Coupon rates are provided by Bloomberg. In this paper, R can be seen as the recovery rate for a default event over the entire bond's maturity because it is extremely rare that a catastrophe bond will default twice. (In my sample, there is no bond default twice.)

Additionally, I use the average high-risk corporate bond yield to control investors' time-varying perception of high-risk bond yield. Based on the nature of the risk level, most catastrophe bonds have a rating of BB or lower. The rating here only reflects the risk of the natural disaster that is covered by the bond. It is reasonable to include the average high-risk bond yield as a control variable because lots of catastrophe bonds are similar to those high-risk corporate coupon bonds in terms of risk rating. If investors ask for a higher yield on high-risk corporate bonds, they probably will also ask for a higher yield on catastrophe bonds.

I use the risk event dummy variable to capture this short term predictability and the remaining effect of the potential disaster event. The risk event dummy variable equals one if the transaction happened during a disaster event period and is zero otherwise. For hurricane and

windstorm related catastrophe bonds, we can predict those disasters' potential movement relatively accurately a week before the disaster event. Therefore, for hurricane and windstorm related bonds, the disaster event period is one week before the land-falling to one week after the hurricane or windstorm's departure. For other catastrophe bonds, like earthquake, the disaster event period is one week after the event's occurrence.

In this paper, I use both bond size and bid-ask spread to control transaction cost and market demand on each bond. The large size of the bond may imply that the issuer is confident about the potential market demand. Therefore, the transaction cost could be smaller for those larger sized bonds. I regress the annual average number of transactions of a bond on bond 'size, EL, coupon, recovery rate, and climate dummy variable and find that bond size plays a positive and significant role in the number of transactions. Bid-ask spread can capture daily market demand that a bond's size can't. Meanwhile, the bond's trigger mechanism also could affect the bond's yield. Investors of bonds with indemnity trigger mechanism normally need a longer time to estimate the loss from a disaster event because an indemnity mechanism bond determines whether it defaults or not on the actual loss from the disaster event. Hence, investors on those bonds could ask for a higher yield. I use an indemnity dummy to indicate whether a bond has an indemnity trigger or not.

3.3. Summary Statistics

Table 1 presents the main characteristics of the bond sample in this paper. On average, the annual expected loss of climate related bonds is higher than the annual expected loss of non-climate related bonds. Non-climate related bonds, on average, have longer initial years to maturity with an average of 3.71 years. On the surface, non-climate related bonds, on average, have a higher coupon rate/expected loss ratio than climate related bonds.

Table 1: Basic Statistics

Unit of Size is in millions of dollars. Coupon rate, Expected loss, and Attachment rate are all annual rates. Years to maturity is the expected maturity at initial offering time. There are 113 climate related bonds and 80 non-climate related bonds in my sample. In panel C, *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Climate related bond					
	Size	Coupon rate	Expected loss	Attachment rate	Years to Maturity
Mean	\$172.93m	0.066	0.024	0.033	3.11
Std	\$174.01m	0.037	0.017	0.024	0.72
Minimum	\$3m	0.00	0.0021	0.0021	0.98
Maximum	\$1500m	0.17	0.098	0.11	5.01
Panel B: Non-climate related bond					
	Size	Coupon rate	Expected loss	Attachment rate	Years to Maturity
Mean	\$211.43m	0.033	0.011	0.012	3.71
Std	\$133.22m	0.16	0.010	0.0099	0.86
Minimum	\$45m	0.015	0.0001	0.0003	2.02
Maximum	\$550m	0.087	0.037	0.041	5.35
Panel C: Coupon rate and Annual expected loss					
	intercept	Climate dummy	Expected loss	R^2	Number of observation
OLS	0.017***	0.014***	1.50***	0.63	193

But for the same level of expected loss, climate related bonds have a larger coupon rate/expected loss ratio than non-climate related bonds. In some of the literature, researchers consider coupon rates a function of annual expected loss. In my sample, there is a very clear positive correlation between the coupon rate and expected loss. After regressing the coupon rate on both expected loss and climate dummy variable, I found out both coefficients are positive and significant. It seems like the issuers have already priced future climate uncertainty by giving out a higher coupon rate to the bond buyer. Before 2016, there were some pricing agents who offered potential investors a climate change scenario annual expected loss and attachment rate.

After getting all the excess yield from every transaction, I calculated the monthly average excess yield for each bond and got 2284 monthly observations with most of the characteristic variables. The figures below show the size weighted average monthly data from June 2014 to March 2022. Different from the climate related bonds in Figure 1, non-climate related bonds gave us very stable relationships among those three curves along the years except in early 2020 when Covid-19 happened. In Figure 2, bond excess yield increased around 2019 and 2020, which could be related to the massive climate related catastrophe bonds defaults during 2019. Both climate

related bonds and non-climate related bonds have a relatively stable annual expected loss curve.

Figure 1: Climate Related Bonds

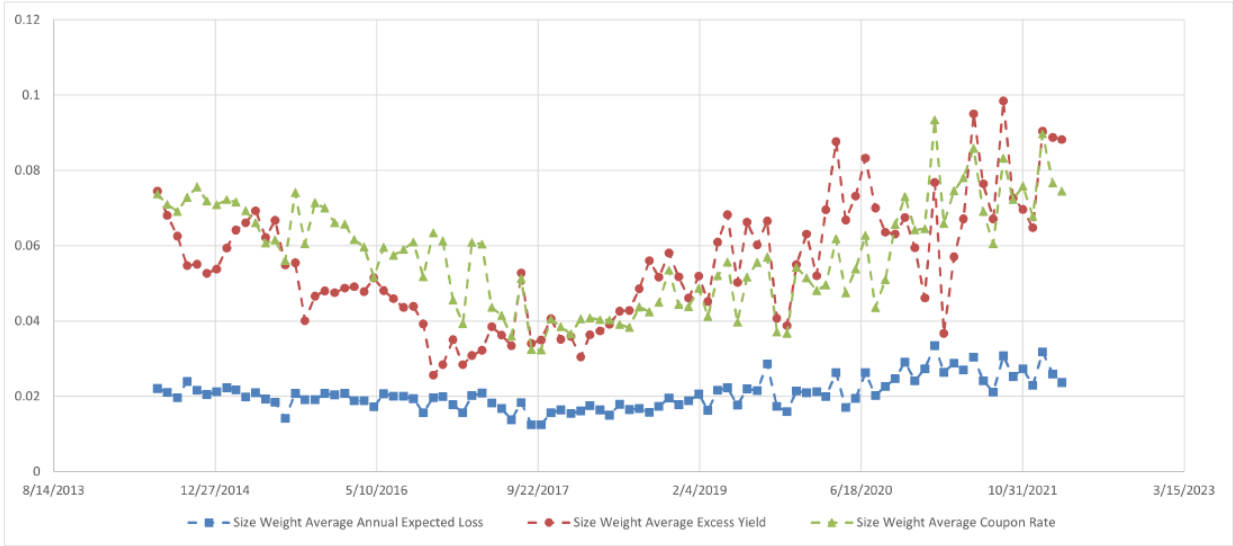
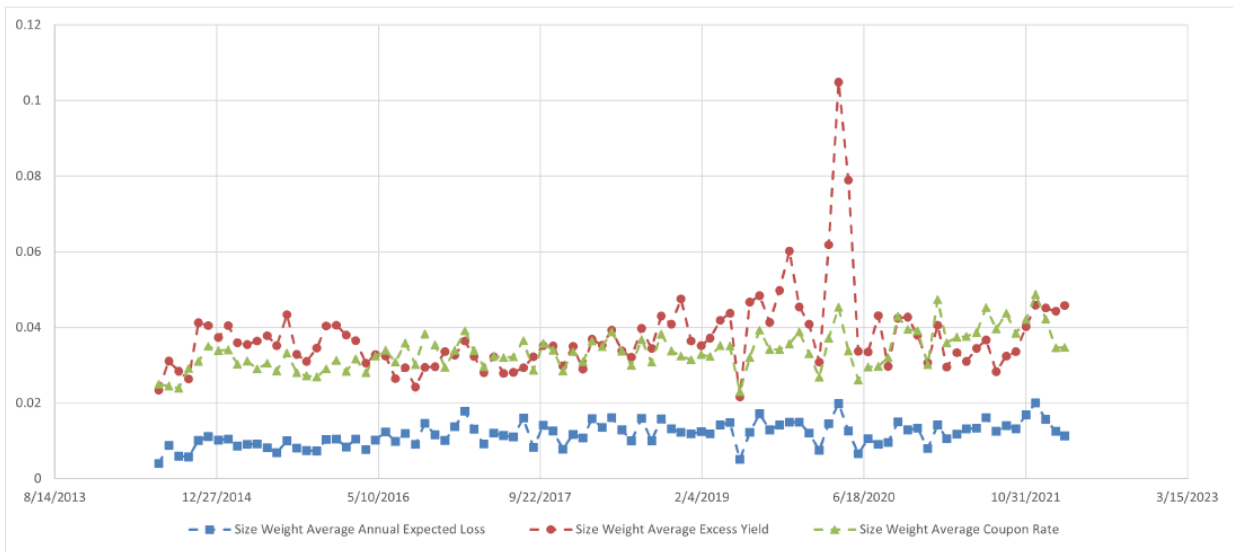


Figure 2: Non-climate Related Bonds



4. MID-TERM CLIMATE RISK PREMIUM

4.1. First Identification

One of the main questions is whether investors will ask for a higher excess yield for climate related bonds when compared with non-climate related bonds for climate change reasons. There are two major differences between climate related bonds and non-climate related bonds other than their attachment rates and expected loss rates which can be controlled in the tests. The first difference is that the natural disaster risk is heavily concentrated in the climate season for climate related bonds, but it can be considered evenly distributed throughout the years for non-climate related bonds. An easy example is that a North American climate related bond, which is bought in January and is going to mature in May, will basically bear no risk at all. The second difference is that investors may have a time changing expectation for future climate related disaster damage loss when compared with non-climate related disasters due to the uncertainty of near-future climate risk concerns. The second difference is the one that this paper wants to test.

I mainly use difference-in-difference approaches to test the second difference. I use a climate dummy variable and a mid-term dummy variable to divide all the sample observations into 4 subgroups which are climate related and short-term bonds, climate related and mid-term bonds, non-climate related short-term bonds and non-climate related mid-term bonds. The key assumption here is that after controlling most of the factors, the difference between mid-term excess yield and short-term excess yield should behave in a similar way among both climate related bonds and non-climate related bonds, if there is no climate change risk concern. I also substitute the mid-term dummy variable to years to maturity in some of the tests which are not typical difference-in-difference tests. Major control variables are annual expected loss of bond (EL), coupon rate, expected recovery rate, and climate season dummy variable. EL can represent the expected level

of the potential disaster event loss, and the expected recovery rate can potentially reflect the variance of the potential disaster event loss. Together EL and R , the expected recovery rate, could capture some investors' preference for risk. Control variable X_{it} also includes risk event dummy variable, risk free rate, high risk yield, size, indemnity dummy, bid-ask spread, and years controls. The regression model is listed below, C_i is the climate dummy variable of bond i , and M_i represents both the mid-term dummy variable and years to maturity of bond i in separate cases.

$$yield_{it} = \alpha + \beta_1 M_{it} + \beta_2 C_i + \beta_3 (M_{it} \times C_i) + \gamma X_{it} + \varepsilon_{it}$$

Results are shown in Table 2. All the excess yields, risk free rates, and high risk yields are monthly averages. In this regression sample, all the transactions occurred before the bonds were triggered to default. Transactions, after the bonds were triggered, will give us extremely low prices which cannot capture investors' risk concern over future disasters through a yield to maturity. As we can see from the regression model, the identification interest is β_3 , the coefficient of $M_{it} \times C_i$. The mid-term dummy variable has its limit since we can't tell the risk premium difference among bonds with more than 1.5 years to maturity. One can imagine that the future climate uncertainty of a climate related bond with 5 years to maturity should be higher than a climate related bond with 2 years to maturity.

Columns (2), (5), and (7) are using EL as the bond's risk level control variable. Ideally, excess yield should have a positive linear relation with EL (Duffie, Singleton 1999). Therefore, EL should be a good risk measurement for bonds. Columns (3), (6), and (8) are using coupon rate as another bond's risk level control variable. A Bond's coupon rate can reflect a certain level of supply and demand relation which might be related to climate change related risk, as we see from

Panel 3 of Table 1. Columns (7) and (8) have bid-ask spread as a control variable to represent the transaction cost. Because TRACE doesn't have all the bid ask records, the number of observations of columns (7) and (8) decreased to 1689. The first row of Table 2 shows all the difference-in-difference results. All the tests use the Mid-term dummy variable to separate those bonds on whether investors can effectively forecast the next and future climate seasons such as hurricane seasons. All the coefficients are positive and statistically significant with meaningful magnitudes ranging from 150 basis points to 210 basis points. This means investors are asking for a higher yield for mid-term climate related bonds than short-term climate related bonds when compared with non-climate related bonds. The second row of Table 2 uses years to maturity instead and shows that climate related bonds with longer years to maturity has an extra yield which non-climate related bonds don't have. Interestingly, all mid-term dummy variables show negative effects with statistical significance, which could be caused by the short-term transaction cost because the catastrophe bond market has low transaction liquidity which probably can't be fully explained by bid-ask spread and the size of the bond. Column (1) and (4) shows a positive coefficient on the Climate dummy but the sign on it changed in all other columns when there are more controls involved. It could be that lots of short-term climate related bonds have a real expected loss rate during their remaining time that is lower than their annual expected loss rate, which is just like the example of seasonality I pointed out above. Results in Table 4 support this idea. Meanwhile, most non-climate bonds have a relatively smaller annual expected loss which is not proportional increase in the excess yield.

Table 2: First Identification

This table shows the main result about the mid-term climate risk premium. Dependent variable is excess yield to maturity. Mid-term dummy variable equals one if the bond will experience at least 2 hurricane seasons and zero otherwise. All the observations are monthly data. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Climate*Mid-term	0.021*** (7.47)	0.015*** (8.06)	0.019*** (10.30)				0.015*** (7.77)	0.020*** (10.57)
Climate*Years to maturity				0.0068*** (6.02)	0.0050*** (6.79)	0.0078*** (10.72)		
Climate	0.0079*** (3.35)	-0.0098*** (-6.05)	-0.015*** (-9.30)	0.0095*** (3.77)	-0.0097*** (-5.45)	-0.017*** (-9.70)	-0.0096*** (-5.85)	-0.015*** (-9.06)
Mid-term	-0.0093*** (-4.39)	-0.0074*** (-5.47)	-0.0056*** (-4.09)				-0.0094*** (-5.26)	-0.011*** (-6.34)
Years to maturity				-0.0031*** (-4.06)	-0.0018*** (-3.58)	-0.00090* (-1.84)	0.00080 (1.47)	0.0028*** (4.92)
EL		1.65*** (44.24)			1.63*** (43.52)		1.67*** (43.74)	
Coupon			0.80*** (44.21)			0.80*** (44.42)		0.81*** (43.94)
R		0.012*** (4.83)	0.0010 (0.42)		0.012*** (4.70)	0.00074 (0.031)	0.010*** (4.02)	-0.0028 (-1.14)
Risk event		0.016*** (9.60)	0.016*** (9.51)		0.017*** (9.62)	0.016*** (9.58)	0.015*** (8.42)	0.014*** (8.25)
High risk yield		0.92*** (8.65)	0.78*** (7.33)		0.93*** (8.68)	0.79*** (7.45)	0.95*** (8.73)	0.80*** (7.36)
Risk free rate		-0.57*** (-4.31)	-0.40*** (-2.99)		-0.63*** (-4.75)	-0.50*** (-3.80)	-0.64*** (-4.69)	-0.50*** (-3.66)
Climate season		0.0011 (0.85)	0.0003 (0.22)		0.00010 (0.095)	-0.00030 (-0.25)	0.0011 (0.83)	0.0010 (0.75)
Climate season * Climate		0.0021 (1.23)	0.0024 (1.44)		0.0038** (2.27)	0.0044*** (2.65)	0.0020 (1.14)	0.0020 (1.18)
Size		-0.0007*** (-3.48)	-0.0006*** (-2.89)		-0.0007*** (-3.38)	-0.0006*** (-2.91)	-0.0007*** (-3.25)	-0.0006*** (-2.89)
Indemnity		0.0027** (2.48)	0.00090 (0.80)		0.0028** (2.51)	0.00040 (0.33)	0.0028** (2.44)	0.00095 (0.084)
Bid-ask spread							5.78*** (6.41)	5.91*** (6.56)
year dummies	No	Yes	Yes	No	Yes	Yes	Yes	Yes
R^2	0.14	0.64	0.64	0.13	0.64	0.65	0.66	0.66
number of observations	2284	1786	1786	2284	1786	1786	1689	1689

R , the expected recovery rate, has a positive and significant loading in columns (2), (5), and (7) when I use EL as risk control. But once I switch EL to the coupon rate, all the statistical significance on R disappears. It could be that the coupon rate itself is a function of both expected loss and expected recovery rate. From the bond issuer's perspective, the coupon rate should reflect the risk of the bond. A 10% increase on R will only cause around 10 basis points increase. It seems like investors prefer more risk at first glance. But the underlying reason for the positive effect of R is that R is correlated with the attachment rate. A low attachment rate means the trigger event needs to be a rarely occurring extreme disaster that causes extreme damage. A high attachment rate means some relatively small disaster could trigger the default of the bond but with a high recovery rate. Therefore, the positive effect from R could be interpreted as investors asking for more yield for a high attachment rate bond. I also did some tests that replace R with AR and found that the results supported the reasoning I mentioned above.

Unsurprisingly, most coefficients on the risk event dummy variable and Indemnity dummy variable are positive and significant. Both size and bid-ask spread play a significant role here with their right direction of signs. Most of the coefficients on the interaction of climate season dummy and climate dummy are positive with no significance.

From Table 2, we can see that there is a clear positive premium for longer term climate related bonds when compared to short-term climate related bonds, but the same situation doesn't apply to non-climate bonds. All the results are still robust even if I apply bond's identity fixed effect regression. But the sample here is unbalanced in term of bonds' identity. What's more, most climate related bonds only face hurricane risk and most non-climate related bonds only face earthquake risk. Therefore, bond's identity fixed effect regression is not the best identification model in this paper.

4.2. Potential Problems

Figure 3: Landfalling

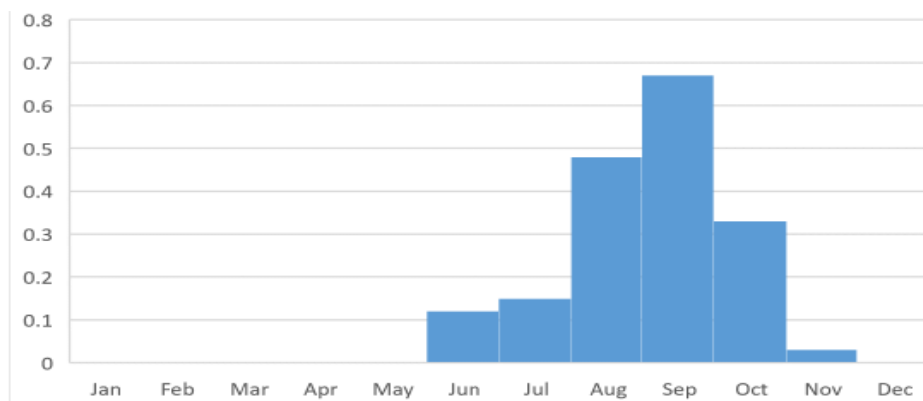


Table 3: Example for Seasonality

	Years to maturity	Number of risky season left
Climate bond1	3.5	4
Climate bond2	1	1
Non-climate bond1	3.5	3.5
Non-climate bond2	1	1

There are some problems that could pose troubles for the tests above. Seasonality of climate related bonds is one of them. As we can see in figure 3, most hurricane related disasters happened during the climate season. And the monthly hazard rates of the disaster are also different.

Here is an example of how seasonality can affect the identification of mid-term risk premium in Table 2. Consider if we have 4 bonds, two climate related bonds, and two non-climate related bonds. All of them have a transaction at the beginning of the climate season and all of them have the same annual expected loss. The only difference among those bonds is years to maturity, all of which are listed in table 3. As we can see from Table 3, the total amount of risk difference between climate related bonds and non-climate related bonds are different even if they have the same difference in years to maturity. Thus, it is reasonable that the yield difference between

climate related bonds should be larger than the yield difference between non-climate related bonds.

One other problem is the term structure issue. In Table 2, the interest of the study is focused on the interaction part of maturity measurement and climate dummy variable. The positive and significant results could be contributed by the term structure difference between climate related bonds and non-climate related bonds if there is a term structure difference.

What's more, different growth rates of local real estate markets could also affect the identification of previous tests. For instance, city A has a 5% annual growth rate and city B has a 2% annual growth rate, which means the same level of disaster would cause more property damage in city A than in city B in the future. The increased yield from longer term climate related bonds could come from the high growth rate of real estate markets in their covered region.

4.3. Robustness Test on Seasonality

Annual expected loss rate can effectively represent the risk level on non-climate related bonds but not the climate related bonds because of the seasonality. Herrmann and Hibbeln (2020) use seasonality adjusted annual expected loss rate showed that there is a seasonality issue in the climate related catastrophe bond market. Their model has some limitations and uses the hazard rate of natural disasters to represent the hazard rate of catastrophe bonds. But because of different trigger mechanisms, the hazard rate of a catastrophe bond is not exactly the hazard rate of its covered natural disaster most of the time.

In this paper, I adapt the idea from Herrmann and Hibbeln (2020) and use seasonality adjusted annual expected loss rate to do the robustness test on seasonality. But I use a different model which addresses the relationship between the hazard rate of catastrophe bonds and the hazard rate of its covered natural disaster. The details of the model are shown in the appendix.

seasonality adjusted annual expected loss rate, denote as $EL_{adjusted}$, is defined as

$$EL_{adjusted} = \frac{remainingrisk}{remainingtime}$$

Intuitively, the seasonality adjusted annual expected loss rate represents the real average annual expected loss for the remaining time of the bond. For example, a north American climate related bond A which will mature in 6 months on June 1st will have a seasonality adjusted annual expected loss rate of zero because its remaining risk is zero. On the contrary, a north American climate related bond B which will mature in 6 months on December 1st will have a seasonality adjusted expected loss rate twice the size of its original annual expected loss. For non-climate related bonds, their seasonality adjusted annual expected loss rates are the same as their original annual expected loss rates.

I first replace the original annual expected loss rate with seasonality adjusted annual expected loss rate in the identification regression model for Table 2. The results are shown in Table 4. The first row in Table 4 still shows a positive and significant effect of mid-term climate related bonds. But the magnitude of the coefficients drops more than 50% when compared with them in Table 2. Interestingly, coefficients on the climate dummy variable are no longer negative as they are in Table 2. It seems that the seasonality adjustment test in Table 4 supports the idea that a lot of climate related bonds have a seasonality adjusted expected loss rate during their remaining time that is lower than their annual expected loss rate. Most of the other control variables in columns (1) and (3) play similar roles as they do in Table 2. In column (2), the coefficient on Climate × Years to maturity is still positive but no longer significant. The reason could be that after seasonality adjustment, some expected loss rate for short term catastrophe bonds dramatically decreases but those bonds were still traded at a relatively high yield because of the transaction cost.

Table 4: Robustness Tests on Seasonality Part A

Dependent variable is excess yield to maturity. Mid-term dummy variable equals one if the bond will experience at least 2 hurricane seasons and zero otherwise. All the observations are monthly data. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)
Climate*Mid-term	0.0060*** (3.34)		0.0052*** (2.74)
Climate*Years to maturity		0.0004 (0.55)	
Climate	0.0016 (1.04)	0.0041 (0.92)	0.0020 (1.24)
Mid-term	-0.0071*** (-5.27)		-0.0054*** (-3.04)
Years to maturity		-0.0012** (-2.02)	-0.00090 (-1.61)
Season adjusted EL	1.62*** (45.17)	1.63*** (44.70)	1.62*** (43.78)
R	0.012*** (5.14)	0.013*** (5.41)	0.010*** (4.10)
Risk event	0.014*** (8.35)	0.014** (8.44)	0.0036** (2.41)
High risk yield	0.91*** (8.67)	0.91*** (8.64)	0.43*** (6.53)
Risk free rate	-0.78*** (-5.98)	-0.76*** (-5.82)	-0.12 (-0.79)
Climate season	0.00090 (0.73)	0.00020 (0.19)	0.0004 (0.28)
Climate season * Climate	0.00080 (0.45)	0.0016 (0.95)	0.00090 (0.50)
Size	-0.00060*** (-2.91)	-0.00060*** (-2.81)	-0.00050** (-2.52)
Indemnity	0.0037*** (3.41)	0.0042*** (3.78)	0.0043*** (3.82)
Bid-ask spread			3.88*** (4.30)
year dummies	Yes	Yes	Yes
R^2	0.65	0.65	0.66
number of observations	1786	1786	1689

Because after seasonality adjustment, the annual expected loss rate distribution of climate related bonds becomes a little bit more unbalanced when compared with it of non-climate related bonds. Therefore, I also match climate related bonds with non-climate related bonds based on their seasonality adjusted annual expected loss rate and years to maturity to conduct the robustness test here. A climate related bond will be matched with a non-climate related bond if their seasonality adjusted annual expected loss rate difference is in the 5% range and their years to maturity are in the 5% range. One bond can be matched with multiple bonds here. Some overlapping matches will be averaged out. For example, if bond A with an $EL_{adjusted}$ 0.03 and years to maturity of 3.00 has been matched with bond B which has an $EL_{adjusted}$ 0.0315. If the same bond B has an observation with years to maturity of 2.95 and an observation with years to maturity of 3.05, I will use the average of these two years to maturities to match bond A. In the end, I have 1786 matched pairs.

After matching $EL_{adjusted}$ and years to maturity, Ideally, a matched pair should have a similar yield to maturity once other factors get controlled. The yield difference between matched bonds should not be affected by the years to maturity of climate related bonds if investors don't have climate risk concerns for their investments. I denote c as climate related bonds and nc as non-climate related bonds. Then the new regression model is

$$y_c - y_{nc} = \alpha + \beta M_c + \gamma X + \varepsilon$$

Due to the lack of observations, no year dummy variables to be used as control variables are here. A bond with a transaction that happened in 2019 can be matched with a bond with a transaction that happened in 2016. X is the control variable list with coupon difference, expected recovery rate difference, risk event dummy difference, risk free rate difference, size difference,

high risk yield difference, and indemnity dummy difference. The main interest is the coefficient of M_c which can be either a mid-term dummy variable or years to maturity of climate related bonds. As we can see from Table 5, both mid-term dummy variable and years to maturity of climate related bonds still have a positive and significant coefficient in columns (1) and (3). And the magnitudes of all three coefficients are at a similar level as we see in Table 4. The loading on years to maturity in column (3) changed its sign to negative with a small t-value. Coupon difference, rate difference, risk event difference, size difference, and high risk yield difference play the same roles as they did in Table 2. Both seasonality adjustments test still provide effective evidence that mid-term climate related bonds have a higher yield than other groups.

Table 5: Robustness Tests on Seasonality Part B

This table shows OLS regression from both seasonality adjusted annual expected loss rate and years to maturity matched data. Dependent variable is the yield difference between two matched observations which are monthly simple average data. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)
Mid-term	0.0037*** (3.11)		0.0042*** (2.81)
years to maturity		0.00076 (1.45)	-0.00039 (-0.59)
Coupon difference	0.68*** (32.8)	0.68*** (32.65)	0.68*** (32.79)
Risk event difference	0.011*** (9.76)	0.011*** (9.96)	0.0064*** (5.30)
Risk free rate difference	-0.18*** (-4.87)	-0.18*** (-4.71)	-0.18*** (-4.85)
Size difference	-0.0018*** (-4.24)	-0.00073*** (-5.16)	-0.00074*** (-5.24)
Indemnity difference	0.0024*** (3.55)	0.0022*** (3.29)	0.0024*** (3.59)
High risk yield difference	0.53*** (11.05)	0.52*** (10.95)	0.53*** (11.05)
R^2	0.68	0.68	0.68
number of observations	1635	1635	1635

4.4. Robustness Test on Term Structure

In Table 2, the identification of the mid-term climate risk premium can be false if the climate related bond has a more sensitive and positive term structure than a non-climate related bond. I need to show that even after controlling the maturity measurement, the mid-term climate risk premium still exists for climate related bonds. Previous tests use maturity measurement to show the mid-term climate risk premium and assume the loading on annual expected loss is a constant. But from another angle, the climate risk premium could come from investors' perception of the bond's annual expected loss and their perception of risk can change depending on the remaining years to maturity. I divided all the bonds into two groups, climate related bonds and non-climate related bonds. And for each group, I slice the sample into 14 subgroups based on the years to maturity of the observations. Because the time span for each subgroup is a quarter of a year, I assume the years to maturity won't play a role here inside the subgroup. I then perform simple OLS for each subgroup with a simple regression model

$$yield_{it} = \alpha + \beta EL_{adjusted} + v_{it}$$

$EL_{adjusted}$ is the seasonality adjusted annual expected loss for bond i . As we can see from Table 6, the magnitude of the loading on climate related bonds increases as the years to maturity increase and becomes stable when years to maturity is bigger than 1. Meanwhile, the magnitude of the loading on non-climate related bonds is relatively stable in all subgroups except the first group. The high coefficient in the first group matches the spike of non-climate bond's excess yield in Figure 2, which was caused by the Covid-19 pandemic. It seems like investors require a higher compensation to bear the same risk for longer term climate related bonds. Therefore, it is

reasonable to assume that the coefficient of EL for climate related bonds is an increasing function of maturity measurement and takes the form of $\zeta + \beta M_{it} + \eta_{it}$. M here is the maturity measurement which can be either a mid-term dummy variable or years to maturity. After plugging the new form of β_i into the original regression model,

$$yield_{it} = \alpha + \beta_1 M_{it} + \beta_2 C_i + \beta_3 (M_{it} \times C_i) + \gamma X_{it} + \varepsilon_{it}$$

we have the new model as seen below

$$yield_{it} = \alpha + \beta_1 EL_i + \beta_2 M_{it} + \beta_3 (M_{it} \times EL_i) + \gamma X_{it} + \varepsilon_{it}$$

In order to have a robustness check on potential different term structures between climate related bonds and non-climate related bonds, I do separate tests on both climate related bonds group and non-climate related bonds group. Unlike the model in Table 2, this model distinguishes some differences between the effect of maturity measurements and the effect of future climate risk concerns. The interest lies in the coefficient of $M_{it} \times EL_i$. In Table 7, I use seasonality adjusted annual expected loss rate instead of the original annual expected loss rate which is used in table 2. As we can see from Table 7 below, $M_{it} \times EL_i$ plays a positive and significant role for climate related bonds but has a negative effect on non-climate related bonds. All the $EL_{adjusted}$ have a positive and significant coefficient. And the magnitude of those coefficients for both climate related bonds and non-climate bonds are at similar levels. All other independent variables play similar roles as they do in the previous tables. most of the maturity measurements have significant and negative effects on the excess yield. This gives us some evidence that term structure does not

play a favorite role in the identification of mid-term climate risk concern. All the results here support the hypothesis that investors ask for higher compensation for mid-term climate related bonds.

Table 6: Robustness Tests on Term Structure Part A

Simple OLS regressions are conducted for each cell of the panel A of the table below. Dependent variable is yield to maturity and the independent variable is annual adjusted expected loss. The coefficients on annual expected loss of each regression are shown in the cells below. All the observations are monthly data. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

		Climate related bonds						
Years to maturity	0-0.25	0.25-0.5	0.5-0.75	0.75-1	1-1.25	1.25-1.5	1.5-1.75	
Coefficient on EL	1.34***	1.39***	1.52***	1.52***	1.94***	2.35***	1.96***	
Years to maturity	1.75-2	2-2.25	2.25-2.5	2.5-2.75	2.75-3	3-3.5	3.5+	
Coefficient on EL	2.43***	2.32***	2.48***	1.98***	1.71***	2.06***	1.62***	
		Non-climate related bonds						
Years to maturity	0-0.25	0.25-0.5	0.5-0.75	0.75-1	1-1.25	1.25-1.5	1.5-1.75	
Coefficient on EL	3.29***	1.27***	1.32***	1.31***	1.15***	1.34***	1.31***	
Years to maturity	1.75-2	2-2.25	2.25-2.5	2.5-2.75	2.75-3	3-3.5	3.5+	
Coefficient on EL	1.23***	1.32***	1.40***	1.30***	1.28***	1.30***	1.27***	

Table 7: Robustness Tests on Term Structure Part B

OLS regressions are conducted here. Dependent variables for all columns are excess yield to maturity. All columns use seasonality adjusted annual expected loss rate. All the observations are monthly data. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	Climate related bond			Non-climate related bond		
	(1)	(2)	(3)	(4)	(5)	(6)
Mid-term * EL		0.33*** (3.46)	0.35*** (3.48)		-0.24** (-2.12)	-0.23** (-2.11)
Years to maturity * EL	0.080** (2.05)			-0.13*** (-2.57)		
Season adjusted EL	1.53*** (18.84)	1.44*** (19.20)	1.46*** (18.13)	1.61*** (14.96)	1.58*** (16.34)	1.60*** (16.8)
Mid-term		-0.0051** (-2.44)	-0.0021 (-0.77)		-0.0046*** (-2.85)	-0.0051** (-2.55)
Years to maturity	-0.0025** (-2.54)		-0.0026** (-2.42)	-0.00060 (-1.03)		0.000055 (0.10)
Risk event	0.014*** (6.49)	0.013** (6.40)	0.012*** (5.41)	0.0081** (1.98)	0.0076* (1.91)	0.0080** (2.04)
Risk free rate	-0.70*** (-3.32)	-0.73*** (-3.52)	-0.75*** (-3.39)	-0.90*** (-6.20)	-0.87*** (-6.03)	-0.86*** (-6.09)
R	0.018*** (4.27)	0.018*** (4.28)	0.017*** (3.99)	0.0050 (1.48)	0.0045 (1.37)	0.0041 (1.27)
High risk yield	0.80*** (4.63)	0.80*** (4.65)	0.81*** (4.57)	1.00*** (8.81)	0.99*** (8.84)	1.01*** (9.17)
Size	-0.0011*** (-4.18)	-0.0011*** (-4.19)	-0.0011*** (-3.93)	0.0013*** (3.57)	0.0011*** (2.98)	0.0011*** (3.10)
Climate season	0.0016 (1.17)	0.0018 (1.32)	0.0012 (0.81)	-0.00040 (-0.35)	0.00060 (0.63)	0.00050 (0.52)
Indemnity	0.0093*** (4.47)	0.0084*** (4.13)	0.0097*** (4.53)	0.00040 (0.30)	-0.00030 (-0.28)	-0.00050 (-0.39)
Bid-ask spread			4.18*** (2.66)			4.82*** (5.43)
Years dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.65	0.65	0.66	0.56	0.57	0.59
number of observations	966	966	899	790	790	790

4.5. Robustness Test on Housing Price

In my data sample, 150 catastrophe bonds have specific covered regions such as Florida, Texas, California, Louisiana, and New Madrid which includes Alabama, Arkansas, Illinois, Indiana, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Ohio, Tennessee, and Wisconsin. Others cover the total United States or other countries. As for now, I have real estate data in the United States from Zillow which provides a state level monthly Zillow Home Value Index (ZHVI) to reflect the monthly price movement of regional real estate markets. I use the state level 3 years moving average monthly growth rate of the Zillow Home Value Index to represent the expected future growth rate of the property market in those states. Then I get the annual growth rate from the 3 years moving average monthly growth rate and use it as an independent variable to control the property market growth rate. If a catastrophe bond covers more than one state, I will use the simple average of their growth rates to represent the growth rate for the covered region of that bond. I exclude bonds that cover the whole United States because the average Zillow Home Value Index growth rate of the United States will be highly biased against bonds' real covered regions. For example, a hurricane bond covers the whole United States despite that most hurricanes will landfall on the coast region only. As a result, the property price growth rate of the United States can't represent those regions effectively.

The results are shown in Table 8. As we can see from Table 8, after controlling the local property price growth, coefficients in both the first row and second row are still positive and significant with similar magnitudes as in Table 2. Not surprisingly, the local ZHVI growth rate also plays a positive and statistically significant role here on a relatively small scale. It is reasonable for investors to ask for a higher yield in those high growth regions because increased property value will make potential default from a disaster easier. Coefficients on both EL and

Table 8: Robustness Tests on Housing Price

OLS regressions are conducted here. Dependent variables for all columns are excess yield to maturity. All the observations are monthly data. Column (1) and (2) use mid-term dummy variable to represent the M_i , and column (3) and (4) use years to maturity to represent the M_i . Column (5) use seasonality adjusted annual expected loss rate instead of original annual expected loss rate. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)	(5)
Climate*Mid-term	0.014*** (6.31)	0.019*** (8.30)			0.0060*** (2.69)
Climate*Years to maturity			0.0056*** (5.34)	0.0082*** (7.91)	
Climate	-0.0083*** (-4.16)	-0.017*** (-8.22)	-0.0090*** (-3.94)	-0.019*** (-8.18)	0.0019 (0.98)
Mid-term	-0.0080*** (-4.42)	-0.0055*** (-3.01)			-0.0064*** (-2.95)
Years to maturity			-0.0023*** (-2.78)	-0.00080 (-1.02)	-0.00040 (-0.49)
ZHVI growth rate	0.16*** (6.17)	0.10*** (3.89)	0.15*** (5.87)	0.090*** (3.40)	0.19*** (7.38)
EL	1.62*** (36.82)		1.60*** (35.98)		1.62*** (38.97)
Coupon		0.78*** (36.03)		0.77*** (35.92)	
R	0.011*** (3.68)	0.0027 (0.90)	0.011*** (3.60)	0.0016 (0.52)	0.015*** (5.09)
Risk event	0.021*** (9.78)	0.019*** (8.68)	0.022*** (9.86)	0.020*** (9.00)	0.019*** (9.20)
High risk yield	0.82*** (6.47)	0.73*** (5.67)	0.80*** (6.25)	0.67*** (5.36)	0.72*** (5.87)
Risk free rate	-0.55*** (-3.41)	-0.56*** (-3.40)	-0.58*** (-3.56)	-0.58*** (-3.56)	-0.70*** (-4.49)
Climate season	0.00091 (0.055)	-0.00070 (-0.41)	-0.0012 (-0.69)	-0.0014 (-0.85)	-0.00010 (-0.090)
Climate season * Climate	0.0021 (0.98)	0.0033 (1.55)	0.0039* (1.83)	0.0052** (2.48)	0.00090 (0.43)
Size	-0.0010*** (-4.20)	-0.00090*** (-3.56)	-0.0010*** (-4.06)	-0.00080*** (-3.38)	-0.00080*** (-3.56)
Indemnity	0.0040** (2.49)	0.0062*** (3.85)	0.0038** (2.38)	0.0057*** (3.55)	0.0072*** (4.74)
Bid-ask spread	8.55*** (6.94)	8.88*** (7.12)	8.31*** (6.69)	8.33*** (6.70)	5.40*** (4.54)
Years dummies	Yes	Yes	Yes	Yes	Yes
R^2	0.69	0.68	0.69	0.69	0.71
number of observations	1247	1247	1247	1247	1247

Coupon in all columns are very similar as they are in Table 2. It seems like the catastrophe bond issuers didn't price the local property value growth into the annual expected loss. All other control

variables play similar roles with similar magnitudes as they do in Table 2.

4.6. Implied Hazard Rate

A more direct way to see how investors measure the potential future climate concern is to use the implied hazard rate from the bond. Implied hazard rates are calculated from the pricing model listed in the Appendix. The pricing model I am using from the Appendix assumes that investors are risk neutral. In lots of bond pricing literature, researchers use risk neutral probability to price the bond. Unlike normal corporate bonds, catastrophe bonds directly connect with the natural risk and have little connection to the general security markets. State prices from other markets cannot represent the risk price here for the catastrophe bond market. The catastrophe bond market is a relatively small market and can be seen as an incomplete market that cannot hedge the risk easily. Therefore, without knowing the investor's utility function, I use a simple risk neutral pricing model to deduce the implied hazard rate for those catastrophe bonds. Even with the assumption of risk neutral for investors, those implied hazard rates can still provide us some insight into the mid-term climate risk concern.

Here, implied hazard rates for both climate related bonds and non-climate related bonds are all monthly average values. Similar to the first test, independent variables are the same and the regression model is

$$hazardrate_{it} = \alpha + \beta_1 M_{it} + \beta_2 C_i + \beta_3 (M_{it} \times C_i) + \gamma X_{it} + \varepsilon_{it}$$

Results from Table 9 seem to tell us the similar story as Table 2 and Table 8. Columns (1) and (3) are using the Mid-term dummy variable as maturity measurement whereas columns (2) and (4) are using years to maturity instead.

In both columns (1) and (3), All the coefficients on climate \times Mid-term are positive and significant with a large magnitude respectively. Catastrophe bonds' average hazard rate will increase about 260 to 290 basis points if the bond still has at least 2 climate seasons left. Interestingly, the coefficient on climate \times Years to maturity is positive but not significant. All the control variables present a similar impact on the implied hazard rate just like what they did for excess yield to maturity. In order to control the local property value growth rate, I conduct similar tests as I did in the previous section but use the implied hazard rate as the dependent variable. Columns (3) and (4) are the test results. After controlling the local property value growth rate, the interactions of the Climate dummy variable and maturity measurement in both (3) and (4) are still positive and statistically significant. The coefficients on row 3 and row 4 are still negative, which could be caused by the short term transaction cost. Just like Table 8, the ZHVI growth rate still plays a positive role.

I also use seasonality adjusted implied hazard rates to test the hypotheses. I use two different seasonality adjustment methods here. First, I use 0-1 adjustment which assumes climate related bonds have a zero hazard rate during non-climate season and have a constant hazard rate during the climate season (June 1st to November 30th). Other than 0-1 adjustment, I also use monthly historical hurricane frequency adjustment which is shown in Figure 3 to mimic the seasonality of climate related bonds. Seasonality adjusted implied hazard rate has a strong short term bias. Because I assume zero hazard rate outside the climate season, all the short term transaction costs will be calculated into the climate reason and be identified as hazard rate.

For instance, East Lane Re V 2012 A (windstorm bond) has one transaction with an average price of 101.91 dollars on November 25th, 2015 and will mature on March 16th, 2016. The yield is 2.97% and the riskless rate at that time is about 0.2%. There is no sign that the bond will default

before maturity. Based on the monthly historical hurricane frequency adjustment, this bond only had 5 days under the potential risk. In this transaction, all the excess yield will be counted into those 5 days in the remaining November 2015 and make the seasonality adjusted implied hazard rate increase dramatically. From that transaction, I get the average hazard rate (no seasonality adjustment) which is 0.033, the 0-1 seasonality adjusted hazard rate which is 0.28, and the monthly historical hurricane frequency adjusted hazard rate which is 2.95. This type of bias will dramatically decrease as the years to maturity increase. Therefore, I use observations which have years to maturity larger than 0.2 to conduct the tests, which are shown in Table 10. In Table 10, columns (1) and (2) are using 0-1 seasonality adjustment, and columns (3) and (4) are using monthly historical hurricane frequency seasonality adjustment. As we can see, from Table 10, all coefficients on both Climate \times Mid-term and Climate \times Years to maturity are positive and significant with a similar magnitude as they are in Table 8. But the significance of other control variables is smaller than they are in Table 9. All the tests so far provide us with a clear picture that there is a difference between climate related bonds and non-climate related bonds. And this difference is not driven by the seasonality of climate related bonds, bonds' term structure, and different local property value growth rates.

Table 9: Implied Hazard Rate Part A

OLS regressions are conducted here. Dependent variable is the implied hazard rate of the bond. M is the maturity measurement which can be represented as either mid-term dummy variable or years to maturity. Mid-term dummy variable equals one if the bond will experience at least 2 hurricane seasons and zero otherwise. Column (1) and Column (3) are using Mid-term dummy variable whereas column (2) and column (4) are using years to maturity. All the observations are monthly data. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)
Climate*Mid-term	0.026*** (3.50)		0.029*** (2.87)	
Climate*Years to maturity		0.0019 (0.62)		0.0096** (2.14)
Mid-term	-0.043*** (-7.82)		-0.048*** (-6.02)	
Years to maturity		-0.016*** (-7.98)		-0.027*** (-7.64)
Climate	-0.0099 (-1.50)	-0.0011 (-0.16)	-0.0088 (-0.98)	-0.0095 (-0.94)
EL	0.13 (0.46)	0.35 (1.23)	0.28 (0.78)	0.65* (1.82)
Coupon	1.15*** (8.47)	1.01*** (7.43)	1.07*** (6.17)	0.92*** (5.35)
Expected recovery rate	0.14*** (14.47)	0.15*** (15.40)	0.16*** (12.52)	0.18*** (13.77)
Risk event	0.039*** (5.72)	0.041*** (5.94)	0.056*** (5.85)	0.055*** (5.88)
Risk free rate	-0.90* (-1.69)	-0.82 (-1.54)	-0.55 (-0.77)	-0.59 (-0.84)
Climate season * Climate	-0.014** (-2.05)	-0.0091 (-1.37)	-0.016* (-1.70)	-0.0077 (-0.86)
Climate season	-0.0004 (-0.077)	-0.0068 (-1.35)	-0.00030 (-0.035)	-0.0099 (-1.38)
High risk yield	1.05** (2.46)	1.13*** (2.65)	1.37** (2.46)	1.45*** (2.64)
Size	-0.00090 (-1.11)	-0.00070 (-0.90)	-0.00090 (-0.87)	-0.0012 (-1.13)
Indemnity	0.0012 (0.27)	0.0062 (1.39)	0.0051 (0.73)	0.0021 (0.31)
Bid-ask spread			-1.37 (-0.26)	0.72 (0.13)
ZHVI growth rate			0.086 (0.75)	0.17 (1.50)
year dummy	Yes	Yes	Yes	Yes
R^2	0.29	0.30	0.28	0.28
number of observations	1786	1786	1247	1247

Table 10: Implied Hazard Rate Part B

OLS regressions are conducted here. Dependent variable is the implied hazard rate of the bond. M is the maturity measurement which can be represented as either mid-term dummy variable or years to maturity. Mid-term dummy variable equals one if the bond will experience at least 2 hurricane seasons and zero otherwise. Column (1) and Column (3) are using Mid-term dummy variable whereas column (2) and column (4) are using years to maturity. Column (1) and Column (2) are using 0-1 seasonality adjustment. Column (2) and column (4) are using monthly historical hurricane frequency seasonality adjustment. All the observations are monthly data with years to maturity at least 1 year. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)
Climate*Mid-term	0.032*** (9.12)		0.034*** (10.11)	
Climate*Years to maturity		0.012*** (8.11)		0.012*** (8.74)
Mid-term	-0.015*** (-5.51)		-0.015*** (-5.61)	
Years to maturity		-0.0054*** (-5.62)		-0.0052*** (-5.61)
Climate	-0.018*** (-5.53)	-0.020*** (-5.55)	-0.019*** (-5.95)	-0.020*** (-5.78)
EL	0.80*** (6.14)	0.79** (5.89)	0.87*** (6.86)	0.86** (6.60)
Coupon	0.73*** (11.40)	0.73*** (11.09)	0.73*** (11.67)	1.14*** (17.47)
Expected recovery rate	0.13*** (29.06)	0.14*** (28.87)	0.13*** (29.71)	0.13*** (29.39)
Risk event	0.043*** (13.56)	0.043*** (13.73)	0.039*** (12.78)	0.040*** (4.14)
Risk free rate	0.27 (1.06)	0.19 (0.74)	0.31 (1.29)	0.23 (0.94)
Climate season * Climate	-0.0066** (2.12)	-0.0023 (-0.74)	-0.0086*** (-2.84)	-0.0040 (-1.33)
Climate season	0.0030 (1.28)	0.00080 (0.32)	0.0035 (1.51)	0.0012 (0.54)
High risk yield	0.73*** (3.51)	0.75*** (3.59)	0.83*** (4.10)	0.85*** (4.16)
Size	-0.0013*** (-3.26)	-0.0012*** (-3.17)	-0.0012*** (-3.21)	-0.0012*** (-3.12)
Indemnity	0.011*** (5.22)	0.011*** (5.44)	0.010*** (5.05)	0.011*** (5.24)
year dummy	Yes	Yes	Yes	Yes
R^2	0.61	0.61	0.63	0.62
number of observations	1689	1689	1689	1689

5. AN EVENT STUDY

All the results so far indirectly provide some evidence that investors tend to ask for a higher yield for longer term climate related bonds. But none of the tests explain where this mid-term climate risk premium comes from. There are two ways to classify potential channels of this mid-term climate risk premium: one is the mid-term prediction of the climate situation and the other one is the investor's increased risk aversion towards the uncertainty of the unpredictable mid-term future climate risk.

As of now, climate predictions for longer than a year are not effective. But there are some less than a year prediction on hurricane season. Many agents including NOAA (National Oceanic and Atmospheric Administration) will release annual forecasts at beginning of the year or at the beginning of the hurricane season based on sea surface temperature and other climate factors. Those annual forecasts can be controlled by year dummy controls to some degree. Short term predictions on hurricanes can be explained by the risk event dummy variable. The mid-term climate prediction could be an omitted variable that investors know but I don't have data yet. Therefore, I am looking into the second channel here in this paper. With the absence of effective mid-term prediction of climate disasters and the existence of long-term climate change background, it is reasonable for investors to ask for more compensation for the unpredictable mid-term future climate risk. Unlike non-climate related disasters which have a stable risk level in the future, climate related disasters have a risk level that can change in the future in either direction. This unpredictability of risk could increase an investor's risk aversion level. If the unpredictability is one of the main reasons to explain this extra risk premium, then one can imagine that the unpredictability will increase after some extreme rare climate disasters happened in a short period of time. Some existing evidence points out that professional money managers will overreact to

large climatic disasters (Alok, Kumar, and Wermers 2020). In the case of catastrophe bonds, it is likely that fund managers will also overreact to longer term climate related bonds after a major disaster event. After a series of major costly extreme climate related disaster events, the near future climate risk will be even more unpredictable. Investors could increase their risk aversion level after a series of disaster events.

Table 11: Recent Disaster Events

Disaster	Time period	Damage
Hurricane Harvey	2017 Aug	\$125 <i>billion</i>
Hurricane Irma	2017 Sep	\$77.16 <i>billion</i>
Hurricane Maria	2017 Sep	\$91.61 <i>billion</i>
Typhoon Jebi	2018 Aug-Sep	\$18.6 <i>billion</i>
Hurricane Florence	2018 Sep	\$24.2 <i>billion</i>
Hurricane Michael	2018 Oct	\$25.1 <i>billion</i>
California wildfires	2018 Aug-Nov	\$400 <i>billion</i>

The year 2017 and year 2018 provided us with an experiment field to look into this potential overreaction from investors. Three of the top four costly hurricanes in the history of the United States happened in 2017 from August to September alone. Another two of the top eleven costly hurricanes in the United States happened in 2018 from September to October and the infamous California wildfires also happened during August, 2018. All the major climate related disasters are listed in Table 9 below. Those seven events cost more than 700 billion dollars. There are a total of 23 defaults in my data sample and 22 of them are related to these seven climate events. During the extreme disaster period, investors may ask for a higher mid-term climate risk premium because of the increased unpredictability of mid-term future climate risk. In order to test this hypothesis, I use difference-in-difference-in-difference to conduct an event study that focuses on bond transactions happening from the year 2017 to the year 2018 time period. Here, I use daily transaction data of

bonds that are not involved in those disasters from 2017 to 2018. Bonds' transactions that are not involved in those disasters can reflect investors' perception on those events and be used to test the hypothesis effectively. I use the disaster period dummy variable, denoted as D_{it} , to distinguish all the daily transactions. The disaster period dummy variable equals one if the transaction happened between August 2017 and December 2017 as well as between August 2018 and December 2018, and zero otherwise. Even though in 2017 all the major climate disasters ended in October, the estimation of the damage could take a couple of months to measure. Therefore, it is reasonable to treat those bonds in December 2017 as still under the influence of the disaster. Then the new regression model is

$$yield_{it} = \alpha + \beta_1 M_{it} + \beta_2 C_i + \beta_3 D_{it} + \beta_4 (M_{it} \times C_i) + \beta_5 (M_{it} \times D_{it}) + \beta_6 (C_i \times D_{it}) + \beta_7 (M_{it} \times C_i \times D_{it}) + \gamma X_{it} + \varepsilon_{it}$$

X_{it} includes the climate season dummy variable, the intersection between climate dummy variable and climate season dummy variable, annual expected loss rate, coupon rate, expected recovery rate, size, risk free rate, and indemnity dummy variable. The interests will be both β_4 and β_7 . Results are shown in Table 12. From Table 12, we can see that the coefficients for both the first row and the second row are positive and significant with meaningful magnitudes. And $M_{it} \times C_i$ still plays a positive and significant role here but with a smaller scale when compared to previous tests from Table 2. Interestingly, $D_{it} \times C_i$ has a negative and significant effect on the yield. It means short-term climate related bonds will have a relatively low yield during the disaster period. Short-term climate related bonds only have at most three months climate related risk before maturity, whereas the non-climate related bonds counterpart may have more total risk before the maturity. So it is reasonable to see a negative coefficient on $D_{it} \times C_i$. Disaster periods also have a positive

and significant impact on the yield, which could reflect the irrationality of investors in non-climate related bonds since the risk of non-climate related risk is unchanged. Most control variables play similar roles as they do in previous tests. One interesting exception is the intersection between climate season and climate dummy variable now has a positive and significant effect on excess yield to maturity. The results are still robust when I use seasonality adjusted annual expected loss rate instead.

Table 12: Tests for The Event Study

OLS regressions are conducted here. Dependent variable is yield to maturity of the bond. Column (1) and (2) use annual expected loss rate as one of the control variables. Column (3) and (4) use seasonality adjusted annual expected loss as one of the control variables. Mid-term dummy variable equals one if the bond will experience at least 2 hurricane seasons and zero otherwise. All the observations are daily data. t-Statistics are shown in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)
Disaster period * Climate * Mid-term	0.0075** (2.48)		0.0094*** (3.13)	
Disaster period * Climate * Years to maturity		0.0046*** (3.57)		0.0052*** (4.08)
Disaster period * Climate	-0.0077*** (-3.00)	-0.011*** (-4.05)	-0.0096*** (-3.78)	-0.013*** (-4.69)
Climate * Mid-term	0.0078*** (4.32)		0.0048** (2.54)	
Climate * Years to maturity		0.0039*** (5.81)		0.0027*** (3.82)
Disaster period * Mid-term	-0.012*** (-6.35)		-0.0096*** (-3.78)	
Disaster period * Years to maturity		-0.0048*** (-7.15)		-0.0051*** (-7.51)
Disaster period	0.014*** (8.71)	0.015*** (9.70)	0.017*** (10.04)	0.016*** (10.50)
Climate	-0.012*** (-8.39)	-0.013*** (-9.76)	-0.0084*** (-4.95)	-0.010*** (-6.20)
Other controls	Yes	Yes	Yes	Yes
R^2	0.75	0.77	0.75	0.77
number of observations	860	860	860	860

6. DISCUSSION AND CONCLUSION

This paper studies the yield difference between climate related bonds and non-climate related bonds. By using the difference-in-difference approach, I find out that both yield difference and implied hazard rate difference between short-term climate related bonds and mid-term climate related bonds are significantly larger than those differences between short-term non-climate related bonds and mid-term non-climate related bonds. Those differences do not go away after a series of robustness checks. Between the year 2017 and 2018 period, the yield difference increased during the disaster period. Investors' risk aversion towards the uncertainty of the near future climate situation could contribute to these differences.

This paper points out the different treatment of climate related bonds by investors, pays attention to mid-term climate change risk, and indirectly provides some evidence that investors' uncertainty on future climate situation could be the source of this difference between climate related bonds and non-climate related bonds. This paper provides another angle to explore how investors view future climate uncertainty. However, there are some limitations to this paper. First of all, the sample size is relatively small so the bond fixed effect can't be applied effectively in this paper. What's more, there is no direct measurement of climate change in the paper to help identify the results. With more transaction data, we can get a more clear picture of whether the potential climate risk has been priced rationally. Future research can try to identify this mid-term climate risk premium more directly by doing some surveys to understand investors' preferences for climate change. With more years of observations, we can also try to connect whether media coverage of climate change can influence investors' views on this long run risk. With the existence of mid-term climate risk, we can also try to identify what catastrophe bond structure is the best that the issuer can offer.

APPENDICES

APPENDIX A: Seasonality Adjusted Annual Expected Loss Rate

I adapt the concept of seasonality adjusted annual expected loss rate from Herrmann and Hibbeln (2020). I denote seasonality adjusted annual expected loss as $EL_t^{adjusted}$ and original annual expected loss as $EL(1)$. And

$$EL_t^{adjusted} = \frac{RemainingRisk_T}{RemainingTime_T} \approx \frac{\int_t^{t+T} \lambda^d(\tau) d\tau}{\int_t^{t+1} \lambda^d(\tau) d\tau} \cdot \frac{EL(1)}{T} \quad (1)$$

Here $\lambda^d(\tau)$ is the hazard rate of the bond's underlying catastrophe disaster event at time λ . For example, for a hurricane bond which will expire in 5 months on May 1st, its $\int_t^{t+T} \lambda^d(\tau) d\tau$ is 0. Therefore its $EL_t^{adjusted}$ should be 0 instead of the original $EL(1)$. The hazard rate of natural disasters can be obtained from historical disaster event data.

This design of $EL_t^{adjusted}$ fits catastrophe bond with parametric trigger very well. Because the trigger line will get reset after every disaster event. Assume that the hazard rate for catastrophe bond is $\lambda^b(\tau)$ the trigger line is Z , and the probability of a disaster won't trigger the default is $F(Z)$. For a parametric trigger mechanism, we can have

$$\lambda^b(\tau) = \lambda^d(\tau) \cdot (1 - F(Z)) \quad (2)$$

I first define total expected loss per dollar (par value) of a catastrophe bond with years to maturity T at time t as

$$EL_t(T) = (1 - R) \cdot P_t(\text{triggerevent} \leq t + T) \quad (3)$$

where

$$P_t(\text{triggerevent} \leq t + T) = 1 - e^{-\int_t^{t+T} \lambda^b(\tau) d\tau} \quad (4)$$

Here, in equation (1), R is the expected recovery rate, ranging from 0 to 1, if a default event happens. R can be deduced from annual expected loss and attachment rate (details are in section 3.2). I assume the conditional expected recovery rate is predetermined. In equation (3), $P_t(\text{triggerevent} \leq t + T)$ is the probability that the bond will not default between time t to time $t + T$. $\lambda^b(\tau)$ is time-varying hazard rate for catastrophe bond at time τ . For non-climate related bonds, $\lambda^b(\tau)$ is a constant. In this model, I assume that triggered events are exponentially distributed, and bonds can be triggered only once. Then, when $\int_t^{t+T} \lambda^d(\tau) d\tau$ is relatively small,

$$\begin{aligned}
EL_t^{adjusted} &= \frac{RemainingRisk_\tau}{RemainingTime_\tau} = \frac{EL_t(T)}{T} \\
&= \frac{EL_t(T)}{EL(1)} \cdot \frac{EL(1)}{T} \\
&= \frac{(1-R) \cdot (1 - e^{-\int_t^{T+t} \lambda^b(\tau) d\tau})}{(1-R) \cdot (1 - e^{-\int_t^{1+t} \lambda^b(\tau) d\tau})} \cdot \frac{EL(1)}{T} \\
&\approx \frac{\int_t^{T+t} \lambda^b(\tau) d\tau}{\int_t^{1+t} \lambda^b(\tau) d\tau} \cdot \frac{EL(1)}{T} \\
&= \frac{\int_t^{T+t} \lambda^d(\tau) \cdot (1-F(Z)) d\tau}{\int_t^{1+t} \lambda^d(\tau) \cdot (1-F(Z)) d\tau} \cdot \frac{EL(1)}{T} \\
&= \frac{\int_t^{T+t} \lambda^d(\tau) d\tau}{\int_t^{1+t} \lambda^d(\tau) d\tau} \cdot \frac{EL(1)}{T} \tag{5}
\end{aligned}$$

If the bond has a trigger mechanism as indemnity aggregated, the $EL_t^{adjusted}$ can only partly represent the real seasonality adjusted annual expected loss. Bonds with indemnity related trigger mechanism normally will reset their trigger line annually. For example, there is a transaction of a bond with indemnity trigger on June 1st, 2019, the bond has accumulated loss L, and will mature on July 1st, 2021. Assume this bond reset its trigger line on the first day of the year. Ideally, the seasonality adjusted annual expect loss is the $(EL_t(T_1) + EL(1) + EL_t(T_2)) / T$. T_1 is the time period between current June 1st to December 31st, 2019, and T_2 is the time period between January 1st, 2021 to July 1st, 2021. With some assumptions on the distribution of the probability of loss from a disaster event, I can get the approximate number of $EL_t(T_2) / T$. But without knowing the current accumulated loss L, it is hard to get an approximate number of $EL_t(T_1) / T$.

But nevertheless, $EL_t^{adjusted}$ from (1) can still capture the main characteristic of seasonality.

With some assumptions, simulation on bonds with indemnity related triggers shows that the difference between the real seasonality adjusted annual expected loss and $EL_t^{adjusted}$ from (1) are relatively small for most of the case.

APPENDIX B: Average Implied Hazard Rate

I use a typical bond pricing model for 1 dollar face value catastrophe bonds which is

$$p = \sum_{i=1}^N (1 - EL_i) \cdot c_i \cdot e^{-y_i t_i} \quad (6)$$

p is the transaction price, and N is the total number of coupons remaining for the bond after the transaction. c_i is the i th cash flow to the bondholder. When $i < N$, $c_i = coupon + r_i$ and $coupon$ is the predetermined fixed annual coupon rate. r_i is the bond's benchmark risk free rate at bond's i th payment date. When $i = 1$, r_i is the spot rate of the bond's specific benchmark risk free rate at the last coupon date. When $i > 1$, r_i is the forward rate of the bond's specific benchmark risk free rate at future coupon date i . When $i = N$, $c_i = coupon + r_i + 1$.

t_i is the time period between the time investors will receive their i th cash flow and the time they buy the bond. y_i denote the riskless continuously compounded spot yield (from Daily Treasury Yield Curve Rates) for maturity t_i . EL_i is the total expected loss for 1 dollar from the transaction date to the i th coupon date, which is described in equations (3) and (4). This pricing model assumes that investors are risk neutral. With transaction price, risk free rate and forward rate, coupon schedule, and years to maturity, one can back out the implied homogeneous hazard rate of the catastrophe bond. I apply two seasonality adjusted methods in this paper. Firstly, I use 0-1 seasonality adjustment. In 0-1 seasonality adjustment, $\lambda_b(\tau)$ is 0 during the non-climate season and is a constant during the climate season. Other than that, I also use monthly historical

hurricane frequency adjustment, which use $\frac{\int_t^{T+t} \lambda^d(\tau) d\tau}{\int_t^{1+t} \lambda^d(\tau) d\tau}$ instead of T in equation (4).

BIBLIOGRAPHY

BIBLIOGRAPHY

- Darwin Choi, Zhenyu Gao, and Wenxi Jiang. 2020. Attention to Global Warming. *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1112–1145.
- Darrel Duffie, Kenneth J. Singleton. 1999. Modeling Term Structures of Defaultable Bonds. *The Review of Financial Studies*, Volume 12, Issue 4, Pages 687-720.
- Emirhan Ilhan, Philipp Krueger, Zacharias Sautner, Laura T.Starks. Climate Risk Disclosure and Institutional Investors. *Working Paper*.
- Emirhan Ilhan, Zacharias Sautner, and Grigory Vilkov. 2021. Carbon Tail Risk. *The Review of Financial Studies*, Volume 34, Issue 3, Pages 1540-1571.
- Jawad M Addoum, David T Ng, and Ariel Ortiz-Bobea. 2020. Temperature Shocks and Establishment Sales. *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1331–1366.
- Justin Murfin, and Matthew Spiegel. 2020. Is the Risk of Sea Level Rise Capitalized in Residential Real Estate? *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1217–1255.
- Marcello Galeotti, Marc Gurtler, and Christine Winkelvos. 2013. Accuracy of Premium Calculation Models for CAT Bonds - An Empirical Analysis. *The Journal of Risk and Insurance*, Volume 80, No.2, Pages 401–421.
- Marcus Painter. An inconvenient cost: the effects of climate change on municipal bonds. *Working Paper*.
- Markus Baldauf, Lorenzo Garlappi, and Constantine Yannelis. 2020. Does Climate Change Affect Real Estate Prices? Only If You Believe In It. *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1256–1295.
- Markus Herrmann and Martin Hibbeln. Seasonality in Catastrophe bonds and Market-Implied Arrival Frequencies. *Working Paper*.
- Markus Herrmann and Martin Hibbeln. Trading and Liquidity in the Catastrophe Bond Market. *Working paper*.
- Michael Barnett, William Brock, and Lars Peter Hansen. 2020. Pricing Uncertainty Induced by Climate Change. *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1024–1066.
- Philipp Krueger, Zacharias Sautner, and Laura T Starks. 2020. The Importance of Climate Risks for Institutional Investors. *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1067–1111.

- Robert F Engle, Stefano Giglio, Bryan Kelly, Heebum Lee, and Johannes Stroebe. 2020. Hedging Climate Change News. *The Review of Financial Studies*, Volume 33, Issue 3, Pages 1184–1216.
- Samuel H.Cox, and Hal W.Pedersen. 2000. Catastrophe Risk Bonds. *North American Actuarial Journal*, Volume 4, Issue 4, Pages 56–82.
- Olivier Dessaint, and Adrien Matray. 2017. Do managers overreact to salient risks? Evidence from hurricane strikes. *Journal of Financial Economics*, Volume 126, Pages 97-121.
- Shashwat Alok, Nitin Kumar, and Russ Wermers. 2020. Do Fund Managers Misestimate Climatic Disaster Risk. *The Review of Financial Studies*, Volume 33, Issue 3, March 2020, Pages 1146–1183.
- Silvia Amaro. 2021. Climate change is not priced into markets, but its effect could be substantial, experts say. *News, CNBC*.
- Simone Beer, and Alexander Braun. Market-Consistent Valuation of Natural Catastrophe Risk. *Working paper*.
- Stefano Giglio, Bryan Kelly, and Johannes Stroebe. Climate Finance. *Working paper*.
- Zacharias Sautner, Laurence van Lent, Grigory Vilkov, and Ruishen Zhang. Pricing Climate Change Exposure. *Working paper*.
- Zonggang Ma, and Chaoqun Ma. 2001. Pricing Catastrophe Risk Bonds: A Mixed Approximation Method. *Insurance: Mathematics and Economics*, Volume 52, December 2012, Pages 243–254.

CHAPTER 2. Long Distance? Not That Bad: Venture Capitalists' Geographic Coverage and
Their Investments

1. INTRODUCTION

For decades, because of low monitoring cost, many venture capitalists (VCs) abided by the famous "20 minutes rule", which states that a company that is beyond a twenty minute driving distance should not be funded by the venture capital fund. Meanwhile, in the finance literature, long-distance investment is always associated with high costs. What's more, there are many benefits for investors to make a nearby investment. For example, Degryse and Ongena (2005) find that loan rates decrease as the distance between the firm and the lending bank decreases.

Additionally, Tian (2009) points out that investing in nearby companies can help reduce the stage financing cost. Also, Malloy (2005) provides evidence that geographically proximate analyses are more accurate than other analyses. What's more, Hollander and Verriest (2016) find that, upon inception, contracts tend to be more restrictive when firms seek loans from remote lenders. Normally, the longer the distance between the investment and its investor, the higher the cost for the investor to collect enough information on the investment. In the venture capital industry, distance is related to the monitoring cost, and monitoring from VCs often plays an important role in terms of nurturing their portfolio companies. Lerner (1995) finds that distance to the firm is an important determinant of the board membership of venture capitalists. Bernstein, Giroud, and Townsend (2016) show that monitoring from venture capital funds can help portfolio companies increase their innovation and the likelihood of a successful exit.

It seems like there are lots of benefits for investors to focus on their nearby potential investments. But these benefits don't mean local investments are always better than remote investments. Hochberg and Rauh (2013), for example, find that there is a home bias associated with under-performance in private equity. Just like a coin, there are always two sides. The downside of investing in a remote company is the informational disadvantage, which increases

monitoring costs. But on the other hand, the benefit of investing in faraway areas is that a fund's investment selection pool will increase so that the fund's optimal investment portfolio can be improved. Therefore, there is a trade-off in the fund's strategy of investment selection. Different types of funds with their own characteristics and abilities should find their own equilibrium to balance the trade-off between nearby and faraway investments. This paper is going to focus on the necessity and benefit of expanding investment territory geographically.

We first provide evidence that venture capital funds that can't access enough high-quality investment opportunities nearby have relatively high incentives to invest far away.

In the venture capital industry, there are often two types of funds: specialist and generalist.

Specialist normally focuses on certain industries, whereas a generalist diversifies its investments into many different industries. For example, a fund that invests most of its capital in Biotechnology will be recognized as a specialist. But there is no clear cutoff line to define how many resources a fund needs to pour into a certain industry to be considered a specialist. It is all about the concentration level of the fund's expertise resource when other fund's characteristics are controlled. The fund's expertise concentration level is a key part to help the fund determine its investment strategy. In this paper, we use specialist (generalist) as a general term to represent funds with relatively high (low) expertise concentration levels compared with other funds in the same group. Expertise concentration levels can not only affect a funds' investment strategies at the industry level, but also geographically. For example, in a certain city, there are only a finite amount of potential investments in a specific industry, such as computer hardware. A computer hardware specialist located in that city will have a strong incentive to look for suitable investments outside the city and expand its investment territory beyond its location area. Meanwhile, a generalist who can also invest in other industries will have a relatively smaller incentive to go outside, because of

increased monitoring costs. Therefore, specialists, in general, should have a larger geographical investment coverage than generalists do.

As a methodology, we develop a concentration index, a combination of the Herfindahl-Hirschman Index and soft cosine similarity score, to reflect the expertise concentration level of a fund. We find out that funds with a high concentration index will tend to invest further away in general. With a one point increase in concentration index, the average investment distance between the fund and its portfolio companies goes up to 66 miles (concentration index range from 0 to 10). We then look into different categories of funds' investments, such as leading group, non-leading group, first round investments, second round investments, etc. The results are robust in most of the categories except the leading group. It seems that venture capital funds still prefer keeping their investment close to home when they are leading the investment, which requires lots of monitoring. Meanwhile, we show that the effect of funds' concentration index on their geographic expansion decreases when they are located in California, which has a large enough investment opportunity for venture capital funds to choose where to invest. Additionally, our results are robust when we use other geographic coverage measurements. Our results suggest that in order to find suitable investment, specialists, funds with high expertise concentration levels, will go further away than generalists, funds with low expertise concentration levels.

We also test whether expanding investment territory can help venture capital funds improve their exit performance. We use the fund's excess exit rates to represent the fund's exit performance. We find out that funds with a larger average investment distance (or other geographic coverage measurements), in general, perform better than funds with relatively smaller average investment distance in terms of excess IPO rate and excess fail rate. As the average investment distance increases by 1000 miles, the fund's IPO performance will increase about 9%, and the

fund's fail rate will decrease 7%. The main out-performance comes from the venture capital fund's first round investments. The results are robust for both California funds and non-California funds. The effect of a fund's average investment distance on the fund's exit performance are consistent across all funds' characteristic spectrum such as the fund's size, the fund's experience, and the fund's expertise concentration level. What's more, we show that funds with relatively a high average investment distance outperform funds with a low average investment distance in both faraway investments and nearby investments. These results support the idea that expanding a fund's investment selection pool can help funds improve their investment portfolio as a whole. Meanwhile, the same tests also show that funds with a high concentration index outperform funds with a low concentration index in terms of IPO and fail rate in general. Last but not the least, we find that inside the fund itself, the fund's faraway investments outperform its nearby investments in terms of both IPO exits and fail exits. And this result can provide useful information to help the venture capitalists determine their own geographic investment strategy. Our results imply that venture capitalists can benefit from long distance investment.

Our paper contributes to literature related to funds' investment selection strategy in terms of geographic distance. We find out specialists in general focus more on faraway investment than generalists. On top of that, our paper contributes to the portfolio investment theory and points out the importance of expanding venture capital funds' investment pools. We find that funds, no matter specialist or generalist, with larger investment geographic coverage tend to have a better IPO and fail exits. We also contribute to the literature related to investor's home bias. We find out that in general, a fund's faraway investments outperform its nearby investment in terms of excess IPO rate and excess fail rate. Last but not least, our findings contribute to the informational advantage-related literature. We find that funds with a high concentration index outperform funds with a low

concentration index.

2. DATA DESCRIPTION

2.1. Data

Our data comes from the Thomson One venture capital database, which upgraded to Refinitiv Workspace after 2019. This database is widely used by many researchers, such as Bernstein, Giroud, and Townsend (2016). The database has information on both venture capital-backed companies and venture capital funds. For companies backed by venture capital, we can learn their basic information such as location, industry category, business description, exit status etc. Additionally, the database offers companies' financing information including number of rounds, participating fund names, and time of the investment rounds. For venture capital funds, the database tells us their basic information such as fund name, vintage year, size, location, etc. We also know the fund's investment history, including some details on its portfolio companies and its investment distribution in terms of industry, city-level geographic location, etc. As previous literature mentioned, we can only observe one location for each fund. Even though many funds do have branches, the systematic source bias should be relatively small.

2.2. Concentration Index

We use the Herfindahl-Hirschman Index on industry distribution to represent the concentration level of a fund in terms of its investment industry selection. $HHI = \sum_{i=1}^N w_i^2$. w_i is the weight of industry i in a fund's overall investments inside the United States multiplied by 100. N is the number of the industry in this fund. A relatively high HHI of a fund means that most of the fund's investments are concentrated in a specific industry.

There are 10 industry categories, from ThomsonOne, in our sample. They are biotechnology, communications and media, computer hardware, computer software and service, consumer related,

industrial/energy, internet specific, medical/health, semiconductors/other elects., and other products. Let's take .406 Ventures LLC as an example. The fund has 31 portfolio companies in our sample. All 31 of them are located in the United States. Among these 31 companies, 21 of them are in the Computer Software and Services industry, 7 of them are in Internet Specific industry, 1 of them is in consumer related industry, 6 of them are in internet specific industry, 4 of them are in medical health, and 2 of them are cataloged as Medical/Health industry. The HH index (HHI) for this fund is 5151.

Other than Herfindahl-Hirschman Index, the soft cosine similarity score can reflect the average similarity of the fund's investments. We use pre-trained words (1 million word vectors trained with subword information on Wikipedia 2017) from Mikolov, Grave, Bojanowski, Puhersch, and Joulin (2018) to calculate the soft cosine similarity score between two portfolio companies in a fund based on companies' business description. Then, we calculate the average score in the upper triangle of the similarity matrix of a fund and use this average similarity score to represent the similarity level of this fund. Let's take fund .406 Ventures LLC as an example again. Both Axial Healthcare Inc and Great Horn Inc are invested by fund .406 Ventures LLC, and are categorized as Computer Software and Services industry. But Axial Healthcare Inc is a provider of pain management care solutions, whereas Great Horn Inc offers a cloud security platform. The similarity score between these two is 0.46. Meanwhile the similarity score between Axial Healthcare Inc and Iora Health, Inc., which is also invested by .406 Ventures LLC and categorized as Medical/Health industry and provides care management and coordination of care designed specifically for older adults, is 0.6. This example suggests that the similarity score can help catch fund's investment differences within the same industry and discover the similarity across industries. More details about the similarity score are shown in the appendix.

Both the Herfindahl-Hirschman Index and the similarity score can help represent a fund's investment concentration level to a certain degree. In this paper, we merge both methods to define the Concentration Index which is $(HHI \times SIM)/1000$ and denoted as C . The range of C is from 0 to 10.

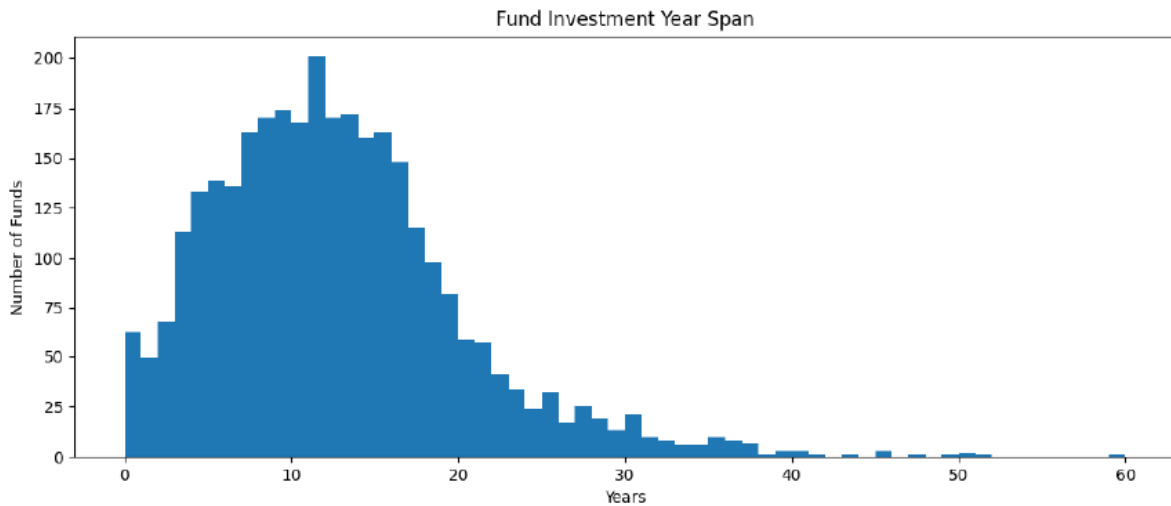
2.3. Control Variables

Some major control variables we use in this paper are fund's size, fund's experience, fund's vintage decade dummy variable, fund's state-level location dummy variables, and fund's investment timing. Size is an important factor that could affect a fund's investment strategy. For example, a larger size fund also means more human resources, which can provide better monitoring and expertise. Only about 65% of the whole fund observations in the Thomson One data base have a size number, but all the fund observations in this paper have a size number. The unit on fund size is a hundred million U.S. dollars in 2019 value. Other than fund size, the fund's experience can also help fund find successive investments and more potential social network connections. Many venture capitalists have raised multiple venture capital funds over the years. The number of funds under their management can reflect their experience in the venture capital industry. Here we use the number of funds under the same venture capitalist as a proxy of experience. For example, Zero Stage Capital III is a venture capital fund managed by Zero Stage Capital Company Inc., which manages 8 funds in total. Therefore the \$Experience\$ variable for Zero Stage Capital III is 8.

A fund's vintage year is also related to the fund's portfolio company selection and their future exit. For example, after the 2000 internet bubble burst, the IPO number per year dropped dramatically. Since the internet era, more and more international companies have gotten investment from U.S. funds. In our sample, around half of the funds have a life span of less than 10 years. And there are a lot of funds that have a life span of more than 20 years. Some funds in

our sample have experienced multiple economic cycles. Instead of fund vintage year control, fund vintage decade control is used in some of our tests. But we also indirectly control our observations at the year level through the benchmark we set for each fund. Figure 4 below shows the distribution of the life span for each fund in our sample.

Figure 4: Fund Investment Year Span Distribution



The location of the fund also plays an important role in the fund's strategy for investment selections. Funds located in California will have less incentive to invest far away because they can find enough high-quality investments in the state. Whereas funds located in relatively small states probably need to explore outside opportunities more often. Meanwhile, each state may have its own economic policy or incentives which can affect venture capital fund's investment decisions.

A fund's investment timing is defined as the average time difference between the date of the fund's first investment in its portfolio companies and the date of the first round of finance its portfolio companies ever received. For example, let's assume a portfolio company received its first-ever investment from investor A in January 2001 and then received a second investment from

investor B in January 2002. Then the investment timing for investor A is 0 years and for investor B is 1 year. It's important to control the fund's investment timing. If a venture capital fund only invests in later round companies, then the average investment distance between it and its investments should be relatively large, because there are not many later round companies in a small region. What's more, funds that invest in more later round companies will have a relatively high IPO exit rate.

2.4. Dependent Variables and Benchmark Exit Rate

For the dependent variable, we use the average geographic distance between the fund and its portfolio companies, fund's faraway investment ratio, and fund's nearby investment ratio to test whether a fund with a high expertise concentration level will go further to find a suitable investment than a fund with a low expertise concentration level when everything else are controlled. In order to see whether investing far away can help funds improve their return, we use excess exit rates to measure the exit performance for each fund. We only include companies that have had at least 4 years to grow since their first round of financing in each fund when we compute these rates. Exit rates include IPO rate, M & A rate, and Fail rate. For each fund-level observation, the IPO rate is defined as the number of companies in the subgroup that went to public, divided by the total number of companies in the subgroup; M & A rate is define as the number of companies that got merged or acquired by other firms in the subgroup, divided by the total number of companies in the subgroup. The fail rate is defined as the number of companies in the subgroup that went into bankruptcy chapter 7, bankruptcy chapter 11, or defunct, divided by the total number of companies in the subgroup.

Table 13: Exit Distribution

Fail includes companies that went bankruptcy and defuncted.

Industry	IPO	M & A	Fail	Active	No. of Company
Biotechnology	0.26	0.25	0.08	0.41	2075
Computer Hardware	0.16	0.34	0.22	0.28	1536
Semiconductors/Other Elect.	0.14	0.36	0.16	0.34	1778
Medical/Health	0.16	0.31	0.14	0.38	3024
Consumer Related	0.14	0.21	0.26	0.39	1588
Communications and Media	0.12	0.40	0.22	0.26	2339
Other Products	0.12	0.25	0.22	0.41	2035
Industrial/Energy	0.13	0.26	0.20	0.42	1510
Computer Software and Services	0.06	0.35	0.09	0.50	7471
Internet Specific	0.05	0.30	0.15	0.49	6066

The excess exit rate is defined as a fund-level observation's exit rate minus that observation's benchmark exit rate. For each year, we calculate different exit rates for all the portfolio investments across different industries. For example, if in 2009, there are 100 internet companies that received investment from venture capitalists, no matter the rounds, and 1 of them make the IPO before 2019, then the IPO rate for the internet industry in 2009 is 1%. Then, we match each year's industry average exit rate to the fund's investment in each year and industry. For example, if fund A made an investment in an internet company in 2009, no matter the rounds, the benchmark exit rates for that investment of 2009 will be matched with the exit rates of the internet industry in 2009. Then we calculate the weighted average on those benchmark exit rates and use it to represent the benchmark for fund A. A fund's benchmark exit rate can reflect the macroeconomic background of that fund.

In Table 13, we organized all of our portfolio companies' exits distribution across all the industry categories. We can see that different industries have quite different exit rates, especially the IPO rate. Biotechnology has the highest IPO rate at 0.26, whereas internet specific industry only has an IPO rate of 0.05. Assume fund A has 20 biotechnology companies, and fund B has

only 10 biotechnology companies and 10 internet specific companies. It's quite normal for fund A to have a higher IPO rate, but we don't know whether its higher IPO rate comes from fund A's expertise or just from the nature of the industry itself. Therefore, we need to control the fund's investment industry composition to show whether expertise can contribute to a higher IPO rate.

2.5. Summary Statistics

Table 14: Fund Type and Its Investment Distribution

Unit of Fund size are in millions of 2019 dollars and unit of Distance is mile. Int distance of a fund is the average distance between its headquarter and its investments' headquarters including international investment. Domestic distance of a fund is the average distance between its headquarter and its investments' headquarters which located in United States. The values in first round, second round, third round, and fourth round and beyond are represent the average percentage of fund's first involvement in its portfolio companies.

Fund Type and Its Investment Distribution			
Investment Stage	Seed Stage	Early Stage	Balanced Stage
First round	0.58	0.45	0.44
Second round	0.18	0.21	0.18
Third round	0.10	0.13	0.13
Fourth round and beyond	0.13	0.20	0.23
Number of Companies (Median)	11	14	17
HHI (Median)	3265	3128	2702
Similarity score (Median)	0.47	0.48	0.46
Fundsize (Median)	32.26	88.82	109.40
Int distance (Median)	644.13	737.54	1025.86
Domestic distance (Median)	523.34	592.02	879.06
Number of funds	181	1543	1038

In this paper, we chose U.S based funds with at least 5 investments, including international investments, with the company's information. If a fund only has 1 investment, then its HH index will be 10,000, which is the highest possible value, and the fund will be determined as a specialist. But when a fund only has 1 investment, how can we know its investment preference? Funds with a small number of investments will make the HH index, similarity score, and exit rate highly biased. Therefore, we have this requirement of 5 investments. Among those funds, we only look

into funds that are categorized as seed stage, early stage, and balanced stage. As we can see from Table 14, the investment round composition among those three categories is quite similar. Other types of funds like buyouts, later stage, and others are quite different from seed stage, early stage, and balanced stage in terms of their exit strategy. Hence, this paper only focuses on seed stage, early stage, and balanced stage venture capital funds. What's more, each fund observation needs to have a fund size available. This gives us 2762 funds remaining in total.

3. FUND'S GEOGRAPHIC COVERAGE

3.1. Concentration Index and Geographic Expansion

A venture capital fund with a high expertise concentration level, which also is referred to as a specialist, will only make investments in areas that match its expertise. Whereas venture capital funds with relatively low expertise concentration levels, also referred to as generalists, tend to diversify their investments in terms of industry. In a fixed area with a limited number of high-quality start-ups, specialists may have a harder time finding enough suitable portfolio companies to invest in than generalists. In order to find enough high-quality investments, specialists need to expand their investment territory geographically when their local investment selection pool is small. For example, Arboretum Ventures 1, a venture capital fund that specializes in the medical and health industry, located in Michigan only has 23% of its investments located in Michigan. Meanwhile, Enterprise Development Fund, L.P., a generalist also located in Michigan, has investments in biotechnology, computer hardware, consumer products, medical and health industry. Enterprise Development Fund, L.P. has 70% of its investment located in Michigan.

When local investment opportunity is relatively limited and if a venture capital fund only focuses on its nearby potential investments, the fund is unlikely to fulfill its all ability to make a profit because of the limited investment efficiency frontier. The best portfolio it can choose will

be dominated by the best portfolio it could have if the fund expands its investment pool geographically when the long-distance cost is smaller than the marginal benefit of geographic expansion. Therefore, there is an incentive for the fund to invest in faraway companies, especially for specialists, when the nearby investment opportunity pool is relatively small. But when the local investment opportunity is relatively large, the marginal benefit from geographic expansion may not be big enough to cover the long-distance cost. In this case, the geographic expansion incentive will be relatively low even for specialists.

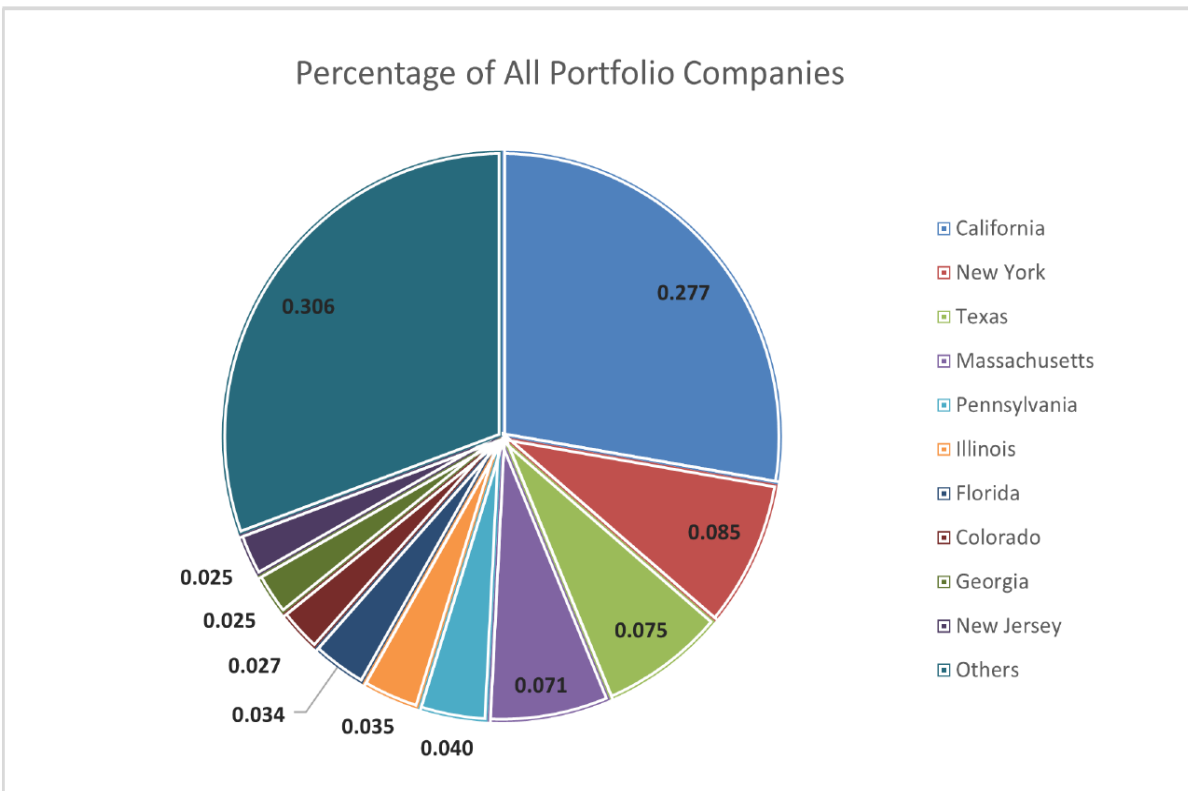
Even though we don't know the magnitude of long distance cost, we should still observe the geographic expansion difference between specialists and generalists, especially for regions that have a relatively limited investment opportunity pool. In terms of empirical analysis, we use a concentration index to represent the fund's expertise concentration level, instead of using a specialist or generalist dummy variable to conduct most of our tests for two reasons. Firstly, there is no strict definition of specialist and generalist. Secondly, even among specialists or generalists, the fund's expertise concentration level should still have an effect on the fund's investment choice geographically. Therefore, we have our first hypothesis that a fund's concentration index, in general, has a positive effect on the fund's geographic expansion and this effect will decrease when the area where venture capital funds are located have a large enough investment opportunity to choose. To test this hypothesis, we need to explain how special California is in our sample first.

3.2. California Funds and California Investments

In our sample, about 27.7% of all the portfolio companies are located in California. New York has 8.5% of all portfolio companies, and Texas has 7.5% of all portfolio companies. The detailed pie chart of the geographic distribution of U.S. portfolio companies is shown in Figure 5. This disproportionate distribution indicates that a venture capital fund located in California may

have an easier time finding suitable investment inside the state than a venture capital fund located in another region of the country.

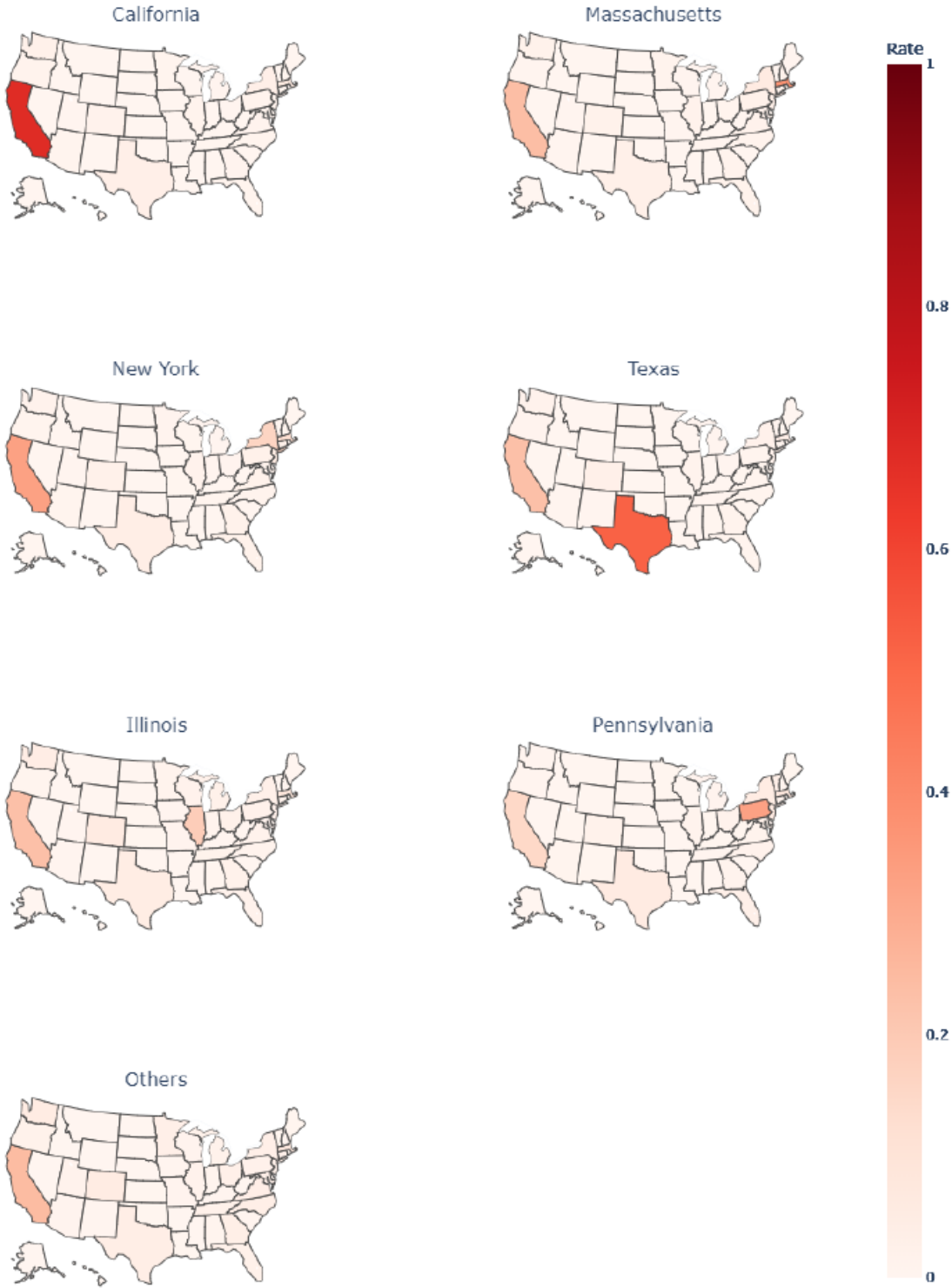
Figure 5: Percentage of All Portfolio Companies



Because of the portfolio companies' disproportional geographic distribution, venture capital fund's portfolio company distribution in the U.S. is not uniformly distributed across the country. Figure 6 shows the fund's investment geographic distribution across the United States. The density of the distribution is shown in different colors on the map. The red color means high density, from 0 to 1, in figure 6. California, Massachusetts, New York, Texas, Illinois, and Pennsylvania are the top 6 states in terms of the number of venture capital funds. As we can see from Figure 6, venture capital funds that are from California most likely will invest in companies located inside California which is different than states such as Massachusetts and others. Outside

of California, 34 out of 49 states have California as one of their top 3 investment destinations in terms of the number of portfolio companies. Both figures 5 and 6 support the idea that funds from regions with a large enough investment opportunity pool have a small incentive to expand their investment territory geographically, whereas funds from regions with relatively limited investment opportunity pools have a relatively large incentive to explore faraway geographically. In our sample, we consider California as the region with a large enough investment opportunity pool. Therefore, the positive effect on the fund's geographic expansion from the fund's concentration index for funds located inside California should be relatively smaller than funds from other regions.

Figure 6: Investments Distribution by States



3.3. Identification

To test our idea on how the fund's concentration index affects the fund's geographic expansion. We regress the fund's geographic expansion level, the fund's average investment distance, on the fund's concentration index while controlling other variables. The identification model is listed below

$$D_i = \alpha + \beta C_i + \gamma X_i + \varepsilon_i$$

Here D_i represents the average investment distance between fund i and its portfolio companies, and the results are shown in Table 15 and 16. Table 15 uses domestic average investment distance (only including the fund's domestic investments) and table 16 uses overall average investment distance (including the fund's international investment). C_i represents the concentration index for fund i . X includes the fund's experience, size, investment timing, vintage decade, and the fund's location. In this section, in addition to testing on the full sample (results listed in column (1)), we also test the other 6 sub-groups which are the leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group. The results are shown in columns (2) to (7) respectively. In this paper, we define a fund as a leader of a portfolio company if the fund invests in the company in the first round and also in most of the company's financing rounds. We also use the definition of the leading group from Bernstein, Giroud, and Townsend (2016), and the results remain the same in this paper. For each fund, a leading group of a fund is defined as a group of investments that is led by the fund. For example, if fund A is the leader for its 10 portfolio companies, then the leading group observation of fund A is from those 10 portfolio companies instead of its all portfolio companies. Similar to

the leading group, for each fund, first round group is defined as a group of investments that are invested by the fund in its first round. Second round group is defined as a group of investments that are first invested by the fund in their second round. Third round group is defined as a group of investments that are first invested by the fund in their third round. Fourth round and beyond group is defined as a group of investments that were first invested by the fund in their fourth round or later round.

As we see from panel A of Table 15, coefficients on C , concentration index, are all positive with similar magnitude ranging from 0.031 to 0.086. Only coefficients from both column (2), the leading group, and column (6), the third round group, are not statistically significant. Because the unit of dependent variable Distance is 1000 miles, a 1 point increase in concentration index (ranging from 0 to 10) could increase the average investment distance by about 66 miles on average in column (1) which represent the full sample test. Results from columns (2) to (7) of both tables show that regardless of whether or not the fund is leading and regardless of the investment round, funds with a high concentration index will look further for suitable investments. Both *Experience* and *Size* play significant positive roles in the fund's average investment distance in both Tables 15 and 16. It seems that larger and more experienced venture capitalists are comfortable with long-distance investments. Probably because they have more human resources to help them manage their investments. As we expected, *InvestmentTiming* has a positive impact on distance. Large value in *InvestmentTiming* means that more first-time later round investments are made by a fund. There are not that many quality companies that deserve later round investment in a certain area. Therefore, the fund needs to go further to find those good investments, and it also supports the idea that the fund will go further if there are not enough options in its nearby area.

Table 15: Concentration Index on Geographic (Domestic Investments)

Dependent variable is fund' average investment distance of US portfolio company only. Unit of distance is 1000 mile. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. D is fund's average investment distance between fund's headquarter and investments' headquarters. C is the concentration index. $Experience$ is number of fund that fund's firm has. $InvestmentTiming$ is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on $Size$ is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Whole United States							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	0.066*** (6.27)	0.031 (1.56)	0.086*** (7.05)	0.060*** (3.12)	0.062** (2.28)	0.042 (0.97)	0.086*** (3.97)
$Experience$	0.0055*** (6.21)	0.0041*** (3.23)	0.0054*** (5.39)	0.0038*** (4.63)	0.0060*** (2.67)	0.0066*** (3.19)	0.0050*** (3.25)
$Size$	0.020*** (8.36)	0.018*** (5.80)	0.021*** (7.82)	0.020*** (6.88)	0.0081** (2.39)	0.0078* (1.92)	0.013*** (3.43)
$InvestmentTiming$	0.053*** (7.46)		0.035*** (4.76)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.29	0.15	0.28	0.24	0.33	0.31	0.32
No. of observations	2753	1217	2484	1702	909	469	927
Panel B: California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	0.031* (1.92)	0.019 (0.55)	0.044** (2.41)	0.013 (0.51)	0.053 (1.32)	-0.012 (-0.19)	0.025 (0.67)
$Experience$	-0.0010 (-0.61)	-0.0027 (-1.02)	-0.00040 (-0.20)	-0.00030 (-0.14)	-0.0042 (-1.52)	0.00046 (0.011)	-0.00030 (-0.082)
$Size$	0.021*** (4.20)	0.029*** (3.63)	0.017*** (3.07)	0.025*** (4.04)	0.0098 (1.24)	0.0010 (1.04)	0.018** (2.25)
$InvestmentTiming$	0.044*** (3.77)		0.033*** (2.70)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.047	0.032	0.028	0.052	0.014	0.16	0.021
No. of observations	990	425	918	608	384	198	350
Panel C: Non-California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	0.089*** (6.45)	0.033 (1.37)	0.11*** (7.04)	0.073*** (3.46)	0.069* (1.89)	0.097* (1.63)	0.13*** (4.61)
$Experience$	0.0082*** (7.56)	0.0060*** (4.13)	0.0082*** (6.69)	0.0070*** (5.22)	0.0072*** (4.21)	0.0080*** (3.65)	0.0065*** (3.75)
$Size$	0.021*** (7.46)	0.016*** (4.93)	0.022*** (7.27)	0.019*** (5.75)	0.0080** (2.15)	0.0082* (1.79)	0.011*** (2.87)
$InvestmentTiming$	0.055*** (6.12)		0.033*** (3.60)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.33	0.30	0.32	0.27	0.34	0.27	0.28
No. of observations	1763	792	1566	1094	525	271	577

Things become more interesting when we look at panel B which only contains funds from California and C which contains all non-California funds of Table 15. In panel B, only columns (1) and (3) have some statistical significance on *C*. And the magnitude of those two columns is very small when compared with them in both panels A and C. Meanwhile in panel C, the situation is totally different. Only column (2) doesn't show statistical significance. What's more, the magnitude of those coefficients on *C* in panel C is much larger than they are in both panels A and B. Meanwhile, the fund's experience is irrelevant in panel B but is important in panel C. The coefficients of *C* and *Experience* in panel B and C show that the incentive for funds to expand their investment territory geographically is relatively small if they are located in California which has enough investment opportunities.

Table 16 uses the fund's average overall investment distance as the dependent variable. We can see that most of the results in table 16 are similar to what they are in table 15 but with a relatively larger magnitude in panel C. The results differ in panels B and C are clearer than they are in table 15. For robustness check, we also use the faraway investment ratio, the number of fund's faraway investments divided by the total number of fund's investment, and nearby investment ratio, the number of fund's nearby investments divided by the total number of fund's investments, as our dependent variables to measure fund's geographic expansion. In this paper, we define an investment is a faraway (nearby) investment if the distance between the investment and the fund is above 300 miles (below 100 miles). Comparing with average investment distance, fund's faraway (nearby) investment ratio can better reflect fund's investment selection strategy geographically. The results are all robust and shown in the Appendix. We find that fund's faraway (nearby) investment ratio is positive (negative) correlated with fund's concentration index in general except California funds.

Table 16: Concentration Index on Geographic (Overall Distance)

Dependent variable is fund' average investment distance of portfolio company including international companies. Unit of distance is 1000 mile. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. D is fund's average investment distance between fund's headquarter and investments' headquarters. C is the concentration index. $Experience$ is number of fund that fund's firm has. $InvestmentTiming$ is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on $Size$ is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Whole United States							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	0.056*** (3.14)	0.048 (1.27)	0.077*** (4.32)	0.044 (1.57)	0.077* (1.88)	-0.0069 (-0.12)	0.096*** (3.39)
$Experience$	0.014*** (9.52)	0.020*** (8.97)	0.012*** (8.94)	0.0055*** (5.72)	0.0066*** (3.05)	0.0061** (2.43)	0.0069*** (3.48)
$Size$	0.037*** (9.06)	0.033*** (5.68)	0.036*** (9.35)	0.039*** (8.04)	0.026*** (5.01)	0.025*** (4.61)	0.025*** (5.89)
$InvestmentTiming$	0.024** (2.01)		0.029*** (2.67)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.18	0.17	0.20	0.19	0.20	0.23	0.22
No. of observations	2762	1225	2489	1709	910	469	929
Panel B: California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	-0.058* (-1.82)	-0.0026 (-0.037)	-0.00050 (-0.016)	-0.066 (-1.37)	0.057 (0.81)	-0.11 (-1.05)	0.028 (0.50)
$Experience$	0.0084*** (2.63)	0.0067 (1.21)	0.0090*** (2.91)	0.0046 (-1.09)	-0.00050 (-0.096)	0.0053 (0.80)	0.0054 (1.14)
$Size$	0.0040 (0.41)	0.0036 (0.22)	0.0083 (0.86)	0.014 (1.16)	0.0069 (0.49)	-0.00440 (-0.28)	0.017 (1.37)
$InvestmentTiming$	0.042* (1.79)		0.049** (2.25)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.041	0.022	0.047	0.022	0.055	0.075	0.032
No. of observations	994	426	921	608	385	198	350
Panel C: Non-California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
C	0.053*** (6.91)	0.049 (1.14)	0.12*** (5.90)	0.095*** (2.79)	0.076* (1.64)	0.051 (0.82)	0.14*** (4.48)
$Experience$	0.018*** (11.05)	0.025*** (10.52)	0.015*** (10.33)	0.010*** (6.65)	0.0072*** (4.66)	0.0077*** (3.30)	0.0074*** (3.78)
$Size$	0.046*** (10.87)	0.039*** (6.69)	0.043*** (11.28)	0.045*** (8.59)	0.030*** (6.23)	0.031*** (6.45)	0.027*** (6.61)
$InvestmentTiming$	0.017 (1.27)		0.020* (1.7)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.29	0.29	0.34	0.29	0.32	0.35	0.29
No. of observations	1768	799	1568	1101	525	271	577

4. GEOGRAPHIC COVERAGE AND PERFORMANCE

4.1. Identification

Tests above show that a fund's expertise concentration level can affect a fund's investment strategy geographically. Increasing investment territory can bring funds long-distance costs, but it could also improve funds' optimal investment portfolios by offering more investment opportunities. It is interesting to see whether funds with a larger average investment distance can outperform funds with a relatively shorter average investment distance in terms of fund's exit performance. We regress different fund's excess exit rates, which are excess IPO rate, excess M & A rate, and excess fail rate, on the fund's average investment distance while controlling fund's concentration index, fund's size, fund's experience, fund's investment timing, fund's location, and fund's vintage decade. The model we use here is

$$E_i = \alpha + \beta_1 D_i + \beta_2 C_i + \gamma X_i + \varepsilon_i$$

Here E_i represents three different excess exit rates for fund i and the results are shown in Table 17. D_i in the model represents the average investment distance of fund i . Other than average investment distance, we also use fund's faraway investment ratio and fund's nearby investment ratio for robustness check. Some papers find some evidence to support that specialists outperform generalists. Kacperczyk, Sialm, and Lu (2005) find that, on average, more concentrated funds perform better after controlling for risk and style differences. Compared with mutual funds, the informational advantage could play an even big role in private equity-like venture capitals. Xi (2009) finds some evidence that supports the positive relationship between specialization and VC's performance. Gompers, Kovner, and Lerner (2009) find a strong positive relationship between the

degree of specialization by individual venture capitalists at a firm and its success. And from the last section, we find evidence that specialists tend to invest further than generalists. Therefore, we need to control the concentration index to avoid the endogenous problem with the model. Our interest in these tests is the coefficients on the fund's geographic investment coverage, such as average investment distance. All the results with the average investment distance are shown in both Tables 17, 18, 19, and 20. Results with fund's faraway (nearby) investment ratio are shown in the appendix.

Table 17 takes a similar structure as Tables 15 and 16. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. We can see that columns (1) to (4) showed strong statistical significance on D with the investment timing control on columns (1) and (3). Very interestingly, the effect from D is strongest when we look into the first round investment group and leading group, which is a subgroup of the first round investment group. The results suggest that the marginal effect of geographic expansion is high for the fund's first round investments. But if we look into the first-time second round investment group, first-time third round investment group, and first-time fourth round and beyond investment group, the significance disappears. There are two reasons that could explain the results. First of all, in order to be considered as an observation in those subgroups, the fund needs to have at least 5 investments in that subgroup. This screening method will cause the funds in the subgroup to usually have a large size and experience with a relatively low concentration index. Therefore, the marginal benefit from geographic expansion for those types of funds are relatively small, and the results will not be as strong as other funds. Secondly, the average investment distance among those later rounds investments is usually larger than it of first round investments and the probability of going to IPO

for later round investments should be much higher than it of first round investments. Conditional on longer average investment distance and high-quality investments, the marginal benefit from geographic expansion should be relatively small as well. Those two reasons may explain why D barely plays a role in column (5) to (7).

Among those first round investments, we can see that geographic expansion can help increase the fund's excess IPO rate and decrease the fund's excess fail rate. Meanwhile, it is worth noting that the excess M & A rate is negatively correlated with D . The negative effect could come from the substitute effect. Instead of going M & A, those portfolio companies might end up going IPO. The negative effect could also come from funds' exit strategy towards their faraway investments. Results from Table 17 show that the negative effect is mainly from funds' faraway investments. On the contrary, funds' nearby investments don't experience this negative effect of the fund's geographic expansion. As for whether venture capital funds have different exit strategies for different types of investments is another interesting topic to explore. But in this paper, excess M & A exit does not represent either positive or negative exit performance. In terms of exit performance, excess IPO rate and excess fail rate are more ideal.

Table 17: Geographic Coverage and Excess Rates

Dependent variables are fund's exit performance including excess IPO rate, excess M & A rate, and excess fail rate Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. *D* is fund's average distance between fund's headquarter and investments' headquarters. *C* is the concentration index. *Experience* is number of fund that fund's firm has. *InvestmentTiming* is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on *Size* is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Excess IPO Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>D</i>	0.014*** (5.06)	0.016*** (4.13)	0.014*** (4.17)	0.018*** (5.13)	-0.00070 (-0.13)	0.012 (1.25)	-0.0018 (-0.21)
<i>C</i>	0.022*** (8.89)	0.011** (2.25)	0.027*** (8.87)	0.014*** (3.53)	0.018** (2.47)	0.024** (2.06)	0.030*** (3.99)
<i>Experience</i>	0.0012*** (5.84)	0.0012*** (4.03)	0.0011*** (4.67)	0.0012*** (4.27)	0.0020 (1.06)	0.0014*** (3.95)	0.0069*** (2.74)
<i>Size</i>	0.0034*** (5.87)	0.0017** (2.26)	0.0038*** (5.68)	0.0019*** (2.66)	0.0036*** (3.94)	-0.00040 (-0.37)	0.0030*** (2.63)
<i>InvestmentTiming</i>	0.021*** (12.48)		0.015*** (8.22)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.18	0.10	0.14	0.083	0.071	0.077	0.067
No. of observations	2762	1225	2489	1709	910	469	929
Panel B: Excess M & A Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>D</i>	-0.027*** (-6.72)	-0.030*** (-5.01)	-0.020*** (-4.12)	-0.028*** (-5.35)	-0.0031 (-0.36)	-0.012 (-0.86)	-0.0035 (-0.35)
<i>C</i>	0.0036 (0.97)	0.026*** (3.31)	-0.0040 (-0.95)	0.018*** (3.00)	-0.0028 (-0.27)	-0.00040 (-0.024)	-0.00090 (-0.11)
<i>Experience</i>	0.00040 (1.23)	0.00080 (1.58)	-0.00020 (-0.51)	0.0012*** (2.75)	-0.00014 (-0.025)	-0.00029 (-0.040)	-0.00050 (-0.82)
<i>Size</i>	0.0031*** (3.56)	0.0019 (1.58)	0.0021** (2.29)	0.0019* (1.77)	0.00030 (0.24)	0.0010 (0.63)	-0.00030 (-0.24)
<i>InvestmentTiming</i>	0.013*** (5.19)		0.012*** (4.65)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.073	0.082	0.062	0.065	0.034	0.077	0.073
No. of observations	2762	1225	2489	1709	910	469	929
Panel C: Excess Fail Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>D</i>	-0.011*** (-4.08)	-0.019*** (-3.83)	-0.0068** (-2.25)	-0.019*** (-4.63)	-0.010 (-1.56)	-0.018* (-1.89)	-0.00050 (-0.072)
<i>C</i>	-0.032 (-1.30)	-0.0066 (-1.00)	-0.0022 (-0.82)	-0.0048 (-1.01)	-0.0090 (-1.17)	0.0017 (0.15)	-0.0039 (-0.69)
<i>Experience</i>	-0.0010*** (-4.83)	-0.0011*** (2.82)	-0.00070*** (-3.03)	-0.0017*** (-4.84)	-0.00060 (-1.32)	-0.0010** (-2.00)	-0.00030 (-0.75)
<i>Size</i>	0.00090 (1.54)	0.00030 (0.27)	0.00060 (0.96)	0.0011 (1.25)	-0.00020 (-0.23)	0.0017 (1.50)	-0.00040 (-0.41)
<i>InvestmentTiming</i>	-0.0072*** (-4.27)		-0.0034** (-2.12)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.11	0.14	0.12	0.10	0.14	0.15	0.15
No. of observations	2762	1225	2489	1709	910	469	929

Table 18: Exit Performance: California VS non-California

Dependent variables are fund's exit performance including excess IPO rate, excess M & A rate, and excess fail rate. Column (1) to (3) are tests on California fund sample, Column (4) to (6) are tests on non-California fund sample. Column (1) and (4) are tests on excess IPO rate, Column(2) and (5) are tests on excess M & A rate, and Column(3) and (6) are tests on excess fail rate. *D* is fund's average distance between fund's headquarter and investments' headquarters. *C* is the concentration index. *Experience* is number of fund that fund's firm has. *InvestmentTiming* is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on *Size* is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)	(5)	(6)
<i>D</i>	0.018*** (3.41)	-0.037*** (-4.51)	-0.013** (-2.07)	0.018*** (3.97)	-0.022*** (-3.14)	-0.024*** (-4.28)
<i>C</i>	0.016*** (2.65)	0.011 (1.17)	-0.0041 (-0.55)	0.012** (2.41)	0.023*** (2.97)	-0.0046 (-0.75)
<i>Experience</i>	0.0024*** (4.63)	0.00060 (0.69)	-0.0022*** (-3.59)	0.00080** (2.19)	0.0013** (2.39)	-0.0014*** (-3.32)
<i>Size</i>	0.00080 (0.50)	0.0043* (1.77)	0.0024 (1.26)	0.0020** (2.43)	0.0012 (0.99)	0.00090 (0.96)
Fund Location				yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.073	0.076	0.065	0.094	0.072	0.11
No. of observations	608	608	608	1101	1101	1101

Table 19: Exit Performance: Size, Experience, and Expertise

Dependent variables are fund's exit performance including excess IPO rate, excess M & A rate, and excess fail rate. All the observations are from fund's first round investment group. We use the same regression as Table 17 and 18 but without the independent variable that we use to do the segmentation. All the number are the coefficients of fund's average investment distance and their t-values in the parentheses. We divide all the observations into bottom 33%, middle 33%, and top 33% in terms of fund's size, experience, and concentration index. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	IPO			M & A			Fail		
	Bottom	Middle	Top	Bottom	Middle	Top	Bottom	Middle	Top
<i>Size</i>	0.016*** (2.93)	0.014*** (3.30)	0.016*** (3.69)	-0.018** (-2.11)	-0.040*** (-5.90)	-0.027*** (-4.82)	-0.014** (-2.24)	-0.014*** (-3.04)	-0.0057* (-1.78)
<i>Experience</i>	0.015*** (3.35)	0.022*** (3.88)	0.010*** (2.56)	-0.022*** (-2.87)	-0.026*** (-3.16)	-0.035*** (-6.86)	-0.014*** (-2.70)	-0.010** (-1.96)	-0.010*** (-2.87)
<i>Expertise</i>	0.022*** (5.10)	0.0064 (1.43)	0.015*** (3.02)	-0.031*** (-4.89)	-0.026*** (-3.76)	-0.028*** (-3.86)	-0.012** (-2.27)	-0.015*** (-3.37)	-0.0059 (-1.46)

One interesting result comes from Table 18, which looks into first round investment of California funds and non-California funds. In Table 18, columns (1) to (3) are tests on the California fund's first round investment group, and columns (4) to (6) are tests on the non-California fund's first round investment group. The results are very similar in both California funds and non-California funds, which is quite different from what we see in Tables 15 and 16. The coefficients on D show that geographic expansion can help improve the fund's IPO exit and decrease the fund's fail exit no matter where the fund is located at. It is important to point out that even for the California fund, which is surrounded by lots of high-quality investment opportunities, the marginal benefit from geographic expansion may still be bigger than the cost that comes with the expansion.

What's more, we divide all the fund observations into three different groups (bottom 30%, middle 33%, and top 33%) in terms of fund's size, fund's experience, and fund's expertise concentration level. We show that the effects of the fund's average investment distance on the fund's exit performance are consistent across all fund's characteristic spectrum. For each group, we perform the same identification regression as previous tests in Tables 17 and 18 but omit the independent variable that we use to separate the sample. All the results are shown in Table 19. We can see that in all cases, the fund's average investment distance has a positive impact on the fund's IPO exit rate and a negative impact on the fund's Fail and M& A exit rates. 25 out of 27 cases are statistically significant. And the magnitudes of those coefficients among each excess exit rate are at similar levels. Even if we divide all the observations into five different groups (bottom 20% to top 20%) in terms of fund's size, fund's experience, and fund's expertise concentration level, the results are consistent in most cases. Table 19 enhances the idea that venture capital funds can improve their investment portfolio by enlarging their coverage geographically.

Table 20: Exit Performance (Among Funds): Faraway VS Nearby

Dependent variables are fund's exit performance including excess IPO rate, excess M & A rate, and excess fail rate. All the observations are from fund's first round investment group. Column (1) to (3) are tests on fund's nearby group sample, Column (4) to (6) are tests on non-California fund's faraway group sample. Column (1) and (4) are tests on excess IPO rate, Column (2) and (5) are tests on excess M & A rate, and Column (3) and (6) are tests on excess fail rate. D is fund's average distance between fund's headquarter and investments' headquarters. C is the concentration index. $Experience$ is number of fund that fund's firm has. $InvestmentTiming$ is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on $Size$ is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

OLS	(1)	(2)	(3)	(4)	(5)	(6)
D	0.012** (2.11)	0.0014 (0.16)	-0.0050 (-0.78)	0.012*** (3.10)	-0.037*** (-6.96)	-0.0049 (-1.38)
C	0.022*** (5.12)	0.0039 (0.600)	-0.0055 (-1.18)	0.022*** (5.92)	0.0054 (1.03)	-0.0014 (-0.40)
$Experience$	0.0015*** (4.46)	0.0015*** (2.93)	-0.0017*** (-4.70)	0.0011*** (3.85)	0.00010 (0.33)	-0.00080*** (-2.92)
$Size$	0.0021*** (2.57)	0.0011 (0.93)	0.0016* (1.80)	0.0028*** (3.85)	0.0012 (1.13)	0.00060 (0.84)
$InvestmentTiming$	0.014*** (5.11)	0.016*** (3.86)	-0.017*** (-5.50)	0.018*** (7.36)	0.012*** (3.75)	-0.0043** (-1.98)
Fund Location				yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes
R^2	0.11	0.077	0.11	0.11	0.071	0.085
No. of observations	1570	1570	1570	1854	1854	1854

When a venture capital fund expands its investment selection pool, both funds are far away investments and nearby investments should be improved. To test this idea, we look into the fund's faraway investment group and the fund's nearby investment group. Just like the fund's first round investment group we see above, the fund needs to have at least 5 faraway (nearby) investments to be considered a valid observation in a faraway (nearby) sample. In this paper, an investment is a faraway (nearby) investment if the geographic distance between the fund's location and the investment's location is above 300 miles (below 100 miles). As results are shown in Table 20, D does have a positive effect on the fund's IPO exit for both nearby and faraway samples. Interestingly, the magnitude of the coefficients for both columns (1) and (4) are very similar. Results also show a negative relation between D and excess fail rate, which is consistent with the

results from Table 17, 18, and 19.

Interestingly, from Tables 17, 18, and 20, we also see that the concentration index plays a significant positive role in the fund's IPO exit in all cases. This result is consistent with the idea that specialists outperform generalists like Xi (2009), and Gompers, Kovner, and Lerner (2009) suggested. Both the fund's experience and fund's size have a positive impact on the fund's IPO exit. As one can expect, the fund's experience can help the fund decrease the fail exit. But the fund's size has no effect on the excess fail rate. It seems that, without experience or other attributes, size itself cannot prevent or won't help prevent a fund's investments from going bankrupt. For the robustness check, we also use fund's faraway (nearby) investment ratio to proxy fund's geographic investment coverage. Overall results are still robust and shown in the appendix.

4.2. Fund's Faraway and Nearby Investments

Results from Tables 17, 18, 19, and 20 show that funds with larger geographic coverage, in general, will outperform funds that focus more on their nearby investment opportunities. But those results cannot tell us the equilibrium investing strategy for venture capital funds in terms of geographic choice. In this section, we look into whether there is a difference between a fund's faraway investment and its own nearby investment in terms of excess exit rates. Even though we don't have a way to measure the long distance cost and don't know how long distance cost will affect a fund's exit performance, venture capital funds themselves can know whether they are under-investing or over-investing faraway portfolio companies based on the results in this section.

To see the difference between a fund's faraway investments and its nearby investment, we pick funds with at least 5 faraway investments and 5 nearby investments to form the sub-sample. Then we calculate the difference between faraway investments' excess exit rates and nearby investments' excess exit rates for each fund in the sub-sample. Because we use the difference inside

the fund itself, the effect of the fund's characteristics like fund size, fund experience, fund concentration index, and others can be canceled out between faraway and nearby investment groups. The model we use here is

$$DE_i = \alpha + \beta_1 DT_i + \beta_2 FNR_i + \varepsilon_i$$

DE_i is the difference of excess exit rates for fund i , DT_i is the investment timing difference of fund i , and FNR_i is the far-near ratio of fund i . The hypothesis in this section is that, with little long-distance cost, there should be no exit performance difference between the fund's faraway investments and the fund's nearby investments conditional on the fund's geographic investment choice strategy. We use both DT_i and FNR_i to proxy the fund's geographic investment choice strategy. Venture capital funds may have different investment timing for their faraway investment than their nearby investment. For example, some funds may be interested in faraway investments if that investment already has received some investment from other funds. Meanwhile, some funds may prefer nearby investments and focus more resources on their nearby investments. Therefore, we use both DT_i and FNR_i to represent the fund's geographic investment choice strategy. With this hypothesis, we can easily see that our focus should be the α of the model.

Test results are shown in Table 21. Column (1) is the full sample test. Columns (2) and (3) are tests on California funds and non-California funds. Columns (4) and (5) are tests on specialist funds and generalist funds. In this section, a fund will be classified as a specialist (generalist) if it is in the top (bottom) 50% of the sample in terms of concentration index. As we can see in Table 9, all the constants, α , are positive for IPO exit and negative for the fail exit. In panel A, only California funds don't show any significance. It is understandable because there are lots of nearby investment opportunities for California funds to choose. In panel C, all the α are negative with

large statistically significance, even for California funds. Interestingly, there is no statistically significant α in panel B. It could be that M & A is more related to the fund's strategy and can be controlled by the fund to some extent. Table 21 indicates that funds, in general, have a better exit performance for their faraway investments. Even though we don't know the long-distance cost, the exit performance difference still can be a good reference to venture capital funds who know their own long distance cost.

One thing that needs to be addressed here is that all the effects from fund's geographic investment coverage, especially for the non-California group, on the fund's exit performance are not because of non-California funds making investments located in California. When we look into the tests that involve California location controls, those California state dummy variables are not significant at all. What's more, in the Appendix, Table 22 shows that California investment has a 7.5% IPO rate which is ranked at number 14 out of 50 states and has a 12% fail rate which is ranked at 37 out of 50 states in terms of avoiding failure. It is more plausible that the positive effects of fund's geographic investment coverage on a fund's exit performance come from funds' improved investment portfolio.

Table 21: Exit Performance (Inside fund): Faraway VS Nearby

Dependent variables are exit performance difference between fund's faraway investments and nearby investments. It includes excess IPO rate difference, excess M & A rate difference, and excess fail rate difference. Each fund needs at least 5 faraway investments and 5 nearby investments. Column (1) to (5) are tests on full sample, California funds, non-California funds, specialist funds, and generalist funds, respectively. Timing difference is the difference between average timing of fund's faraway investments and fund's nearby investments. Far Near Ratio is the ratio between number of fund's faraway investment and number of fund's nearby investments. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Excess IPO Rate Difference					
OLS	(1)	(2)	(3)	(4)	(5)
Constant	0.018*** (2.76)	0.0087 (0.76)	0.031*** (3.05)	0.018* (1.94)	0.017* (1.67)
Timing Difference	-0.0033 (-0.80)	0.00030 (0.054)	-0.0074 (-1.33)	-0.0027 (-0.52)	-0.0042 (-0.65)
Far Near Ratio	-0.0041 (-1.49)	-0.00070 (-0.081)	-0.0064** (-2.06)	-0.0019 (-0.41)	-0.0046 (-1.29)
R^2	0.0030	0.000	0.011	0.0010	0.0040
No. of observations	1039	543	496	520	519
Panel B: Excess M & A Rate Difference					
OLS	(1)	(2)	(3)	(4)	(5)
Constant	-0.0067 (-0.67)	-0.0035 (-0.21)	-0.0032 (-0.21)	0.0013 (0.083)	-0.012 (-0.93)
Timing Difference	0.023*** (3.70)	0.020** (2.20)	0.025*** (2.94)	0.019** (2.09)	0.027*** (3.16)
Far Near Ratio	0.0027 (0.66)	-0.0033 (-0.26)	0.0025 (0.53)	-0.00070 (-0.090)	0.0042 (0.90)
R^2	0.013	0.0090	0.017	0.0080	0.020
No. of observations	1039	543	496	520	519
Panel C: Excess Fail Rate Difference					
OLS	(1)	(2)	(3)	(4)	(5)
Constant	-0.033*** (-4.56)	-0.040*** (-3.68)	-0.025** (-2.01)	-0.036*** (-3.73)	-0.034*** (-3.10)
Timing Difference	-0.0098** (-2.22)	-0.0040 (-0.66)	-0.015** (-2.32)	-0.0012 (-0.22)	-0.020*** (-2.76)
Far Near Ratio	0.0047 (1.59)	-0.0011*** (1.02)	-0.0086 (-0.83)	-0.0088* (1.79)	0.0036 (0.93)
R^2	0.0070	0.0030	0.013	0.0060	0.016
No. of observations	1039	543	496	520	519

5. DISCUSSION AND CONCLUSION

This paper tries to answer the question of whether venture capital funds expand their investment territory geographically. We first find that funds with high expertise concentration level tend to invest further away than funds with low expertise concentration levels. We show that funds with high expertise concentration levels have more incentives to expand their investment selection

pools geographically when there are not enough high-quality investment opportunities near them. After controlling the fund's expertise concentration level, we find that the fund's exit performance is positively correlated with the fund's geographic coverage even for funds from California which have lots of investment opportunities. We also provide evidence that the effect of fund's average investment distance on fund's exit performance are consistent across all fund's characteristic spectrum. Meanwhile, we also provide evidence that supports the idea that specialists, in general, can outperform generalists. Last but not the least, we show that fund's faraway investments, in general, have better exit performance than their own nearby investments, which can provide some guidance for venture capital funds on whether they should expand their investment territory geographically.

But there is still much room for us to explore. Venture capital fund's performance could come from both their investment selection difference and their investment value-adding difference. This paper mainly focuses on investment selection. How to understand the value-adding ability difference between specialists and generalists is still an open question to explore. In the paper, we use exit performance to proxy return performance. Ideally, venture capitalists should reach a return equilibrium between their nearby investment and faraway investment. But exit performance can have a different equilibrium result than the real return. Therefore, in general, whether venture capitalists as a group reach their investment equilibrium between nearby investment and faraway investment is still unknown.

APPENDICES

APPENDIX A: Soft Cosine Similarity Score

Assume a fund has two portfolio companies A and B. The soft cosine similarity score between company A and B is given by

$$similarity(A, B) = \frac{\sum_{i=0}^N \sum_{j=0}^N s_{ij} \cdot a_i \cdot b_j}{\sqrt{\sum_{i=0}^N \sum_{j=0}^N s_{ij} \cdot a_i \cdot a_j} \cdot \sqrt{\sum_{i=0}^N \sum_{j=0}^N s_{ij} \cdot b_i \cdot b_j}}$$

Here $similarity(A, B)$ is the similarity score between company A and company B. a and b are numerical N dimensional vectors to represent a business description of company A and a business description of company B. N is the total number of different words among both the business description of company A and the business description of company B. s_{ij} is the similarity score between each word i and j among those N words. s_{ij} can have many forms. Here $s_{ij} = \max(0, \cos ine(v_i, v_j))^2$ This soft cosine similarity model is from Charlet and Damnati (2017) v_i is a numerical vector of the word i . In this paper, v_i is from pre-trained word data (1 million word vectors trained with subword information on Wikipedia 2017) from Mikolov, Grave, Bojanowski, Puhersch, and Joulin (2018).

There are two constraints when we calculate the average similarity score for a fund. Firstly, if the length of the business description is shorter than 50 characters, it will be excluded from the similarity calculation. For example, Burk Pumps Inc. has a description of ‘Processes equipment.’ which has a length of 21 characters. Metal Cutting Tools Corp has a description of ‘Metal Cutting Tools Corporation manufactures cutting tools.’ which has a length of 59 characters. Secondly, if there is no business description for that company, the company will be excluded from the similarity calculation. We first get a soft similarity score for every two

available companies in a fund and then calculate the average score in the upper triangle of the similarity matrix of the fund and use this average similarity score to represent the similarity level of this fund.

They are some merits of using soft cosine similarity. It can help funds improve their concentration index ranking if they have companies that can be classified as multiple industries and it can also help decrease the concentration index ranking of funds that focus on consumer-related and other products industry categories.

APPENDIX B: Exit rates states distribution

Table 22: Company Exit Rate By States

There are 77864 portfolio companies used to form Table 10. All the rates in the table are exit rates in each state.

State	IPO	M % A	Fail	State	IPO	M % A	Fail
Alabama	0.055	0.25	0.11	Montana	0.036	0.22	0.084
Alaska	0.087	0.17	0.35	Nebraska	0.041	0.21	0.05
Arizona	0.083	0.21	0.076	Nevada	0.090	0.10	0.076
Arkansas	0.041	0.18	0.082	New Hampshire	0.062	0.30	0.10
California	0.075	0.25	0.12	New Jersey	0.115	0.27	0.12
Colorado	0.065	0.26	0.12	New Mexico	0.031	0.21	0.12
Connecticut	0.091	0.22	0.15	New York	0.061	0.19	0.099
Delaware	0.048	0.097	0.062	North Carolina	0.060	0.24	0.11
Florida	0.075	0.20	0.091	North Dakota	0.089	0.20	0.089
Georgia	0.067	0.26	0.12	Ohio	0.058	0.22	0.098
Hawaii	0.044	0.16	0.11	Oklahoma	0.086	0.19	0.08
Idaho	0.096	0.22	0.032	Oregon	0.063	0.25	0.11
Illinois	0.071	0.24	0.11	Pennsylvania	0.063	0.21	0.091
Indiana	0.063	0.25	0.08	Rhode Island	0.036	0.23	0.15
Iowa	0.060	0.27	0.12	South Carolina	0.058	0.21	0.088
Kansas	0.075	0.24	0.06	South Dakota	0.019	0.12	0.058
Kentucky	0.071	0.23	0.11	Tennessee	0.057	0.20	0.074
Louisiana	0.055	0.22	0.12	Texas	0.078	0.22	0.11
Maine	0.024	0.23	0.14	Utah	0.056	0.23	0.067
Maryland	0.068	0.22	0.11	Vermont	0.023	0.22	0.091
Massachusetts	0.10	0.29	0.13	Virginia	0.059	0.26	0.10
Michigan	0.050	0.22	0.11	Washington	0.072	0.24	0.098
Minnesota	0.11	0.27	0.13	West Virginia	0.056	0.31	0.056
Mississippi	0.090	0.22	0.13	Wisconsin	0.046	0.25	0.084
Missouri	0.057	0.24	0.094	Wyoming	0.067	0.20	0.067

APPENDIX C: Exit rates states distribution

Table 23: Faraway Investment Ratio and Concentration Index

Dependent variable is fund' faraway investment ratio, number of faraway investments over total number of investments, of US portfolio company only. An investment is a faraway investment if it is at least 300 miles away from the fund's location. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. *C* is the concentration index. *Experience* is number of fund that fund's firm has. *InvestmentTiming* is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on *Size* is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Whole United States							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>C</i>	0.035*** (6.01)	0.017 (1.50)	0.040*** (6.19)	0.027*** (3.08)	0.034** (2.43)	0.019 (0.92)	0.039*** (3.57)
<i>Experience</i>	0.0023*** (3.26)	0.0041*** (3.23)	0.0025*** (4.83)	0.0024*** (3.91)	0.0014* (1.86)	0.0017* (1.92)	0.0018** (2.37)
<i>Size</i>	0.0083*** (4.94)	0.018*** (5.80)	0.0097*** (7.00)	0.010*** (6.52)	0.0057*** (3.24)	0.0021 (1.08)	0.0042** (2.47)
<i>InvestmentTiming</i>	0.035*** (8.86)		0.024*** (6.10)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.22	0.17	0.22	0.18	0.21	0.24	0.25
No. of observations	2753	1217	2484	1702	909	469	927
Panel B: California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>C</i>	0.014* (1.65)	0.0051 (0.28)	0.021** (2.33)	0.013 (0.96)	0.052*** (2.59)	-0.0011 (-0.038)	0.019 (1.05)
<i>Experience</i>	-0.0017** (-1.98)	-0.0028** (-2.03)	-0.0015* (-1.74)	-0.0019* (-1.71)	-0.0024* (-1.73)	-0.0014 (-0.73)	-0.0018 (-1.14)
<i>Size</i>	0.0091*** (3.50)	0.011*** (2.60)	0.0076*** (2.82)	0.013*** (3.80)	0.0038 (0.95)	0.0099** (2.19)	0.0050 (1.25)
<i>InvestmentTiming</i>	0.028*** (4.60)		0.024*** (3.89)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.054	0.056	0.042	0.066	0.029	0.034	0.020
No. of observations	990	425	918	608	384	198	350
Panel C: Non-California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>C</i>	0.049*** (6.17)	0.017 (1.23)	0.051*** (5.88)	0.035*** (2.95)	0.018 (0.92)	0.043 (1.49)	0.052*** (3.84)
<i>Experience</i>	0.0043*** (7.07)	0.0039*** (4.79)	0.0045*** (6.82)	0.0040*** (5.29)	0.0030*** (3.33)	0.0026** (2.43)	0.0029*** (3.39)
<i>Size</i>	0.012*** (7.23)	0.0082*** (4.39)	0.011*** (6.47)	0.010*** (5.50)	0.0062*** (3.15)	0.0010 (0.46)	0.0041** (2.23)
<i>InvestmentTiming</i>	0.037*** (7.28)		0.022*** (4.44)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.24	0.18	0.23	0.18	0.16	0.12	0.17
No. of observations	1763	792	1566	1094	525	271	577

Table 24: Nearby Investment Ratio and Concentration Index

Dependent variable is fund' nearby investment ratio, number of nearby investments over total number of investments, of US portfolio company only. An investment is a nearby investment if it is at most 100 miles away from the fund's location. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. *C* is the concentration index. *Experience* is number of fund that fund's firm has. *InvestmentTiming* is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on *Size* is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Whole United States							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>C</i>	-0.021*** (-3.63)	-0.0028 (-0.25)	-0.026*** (-4.09)	-0.0084 (-0.94)	-0.030** (-2.27)	-0.018 (-0.89)	-0.027*** (-2.57)
<i>Experience</i>	-0.0019*** (-3.77)	-0.0014* (-1.94)	-0.0018*** (-3.43)	-0.0019*** (-3.12)	-0.00060 (-0.90)	-0.00050 (-0.61)	-0.00070 (-1.00)
<i>Size</i>	-0.0088*** (-6.54)	-0.0064*** (-3.77)	-0.0078*** (-5.73)	-0.0084*** (-5.33)	-0.0033** (-2.02)	-0.00090 (-0.49)	-0.0036** (-2.29)
<i>InvestmentTiming</i>	-0.037*** (-9.39)		-0.027*** (-7.08)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.31	0.26	0.34	0.26	0.37	0.38	0.41
No. of observations	2753	1217	2484	1702	909	469	927
Panel B: California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>C</i>	-0.0070 (-0.79)	0.0015 (0.081)	-0.014 (-1.48)	-0.0003 (-0.021)	-0.047** (-2.30)	0.00090 (0.030)	-0.0097 (-0.50)
<i>Experience</i>	0.0026*** (2.91)	0.0040** (2.76)	0.0024*** (2.63)	0.0026** (2.32)	0.0030** (2.09)	-0.0014 (-0.73)	0.0026 (1.54)
<i>Size</i>	-0.0088*** (-3.24)	-0.011*** (-2.61)	-0.0072*** (-2.59)	-0.012*** (-3.37)	-0.0030 (-0.73)	-0.010** (-2.09)	-0.0043 (-1.01)
<i>InvestmentTiming</i>	-0.30*** (-4.65)		-0.025*** (-3.94)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.055	0.069	0.041	0.064	0.028	0.031	0.023
No. of observations	990	425	918	608	384	198	350
Panel C: Non-California Investments							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>C</i>	-0.030*** (-3.91)	-0.0010 (-3.64)	-0.033*** (-3.97)	-0.012 (-1.02)	-0.015 (-0.89)	-0.045* (-1.76)	-0.038*** (-3.22)
<i>Experience</i>	-0.0037*** (-6.09)	-0.0030*** (-4.79)	-0.0037*** (-5.84)	-0.0036*** (-4.77)	-0.0021*** (-2.65)	-0.0015 (-1.57)	-0.0017** (-2.33)
<i>Size</i>	-0.0095*** (-6.02)	-0.0059*** (-3.19)	-0.0084*** (-5.31)	-0.0081*** (-4.46)	-0.0035** (-2.03)	0.00050 (0.23)	-0.0036** (-2.28)
<i>InvestmentTiming</i>	-0.040*** (-7.86)		-0.026*** (-5.53)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
<i>R</i> ²	0.22	0.15	0.23	0.17	0.18	0.10	0.23
No. of observations	1763	792	1566	1094	525	271	577

APPENDIX D: Fund's Investment Geographic Coverage and Fund's Exit Performance

Table 25: Excess Exit Rates and Faraway Investment Ratio

Dependent variables are fund's exit performance including excess IPO rate, excess M & A rate, and excess fail rate. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. F is fund's faraway investment ratio which follows the definition of each sample group. C is the concentration index. $Experience$ is number of fund that fund's firm has. $InvestmentTiming$ is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on $Size$ is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Excess IPO Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F	0.028*** (3.44)	0.028** (2.11)	0.026*** (2.74)	0.025** (2.28)	-0.0051 (-0.29)	0.021 (0.73)	-0.0075 (-0.33)
C	0.022*** (8.73)	0.011** (2.26)	0.027*** (8.85)	0.014*** (3.51)	0.018** (2.48)	0.024** (2.03)	0.030*** (4.00)
$Experience$	0.0013*** (6.39)	0.0015*** (4.86)	0.0013*** (5.14)	0.0014*** (4.70)	0.00040 (1.07)	0.0021*** (4.03)	0.0014*** (2.75)
$Size$	0.0036*** (6.14)	0.0020** (2.55)	0.0040*** (6.03)	0.0023*** (3.21)	0.0036*** (3.97)	-0.00020 (-0.17)	0.0030*** (2.66)
$InvestmentTiming$	0.020*** (11.96)		0.015*** (8.02)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.17	0.13	0.14	0.071	0.049	0.075	0.067
No. of observations	2762	1225	2489	1709	910	469	929
Panel B: Excess M & A Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F	-0.018 (-1.48)	-0.042** (-2.02)	-0.0080 (-0.60)	-0.043*** (-2.57)	0.037 (1.44)	0.032 (0.80)	-0.021 (-0.77)
C	0.0027 (0.73)	0.026*** (3.21)	-0.0052 (-1.23)	0.018*** (2.99)	-0.0041 (-0.40)	-0.00070 (-0.043)	-0.00040 (-0.048)
$Experience$	0.00078 (0.25)	0.00030 (0.61)	-0.00040 (-1.18)	0.0010** (2.30)	-0.00089 (-0.16)	-0.00020 (-0.21)	-0.00050 (-0.80)
$Size$	0.0023*** (2.64)	0.0014 (1.10)	0.0015 (1.62)	0.0013 (1.21)	-0.000086 (-0.0070)	0.00060 (0.38)	-0.00030 (-0.22)
$InvestmentTiming$	0.013*** (5.06)		0.012*** (4.46)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.059	0.066	0.055	0.053	0.037	0.077	0.073
No. of observations	2762	1225	2489	1709	910	469	929
Panel C: Excess Fail Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F	-0.023*** (-2.83)	-0.039** (-2.26)	-0.016* (-1.83)	-0.034*** (-2.59)	-0.030 (-1.52)	-0.071*** (-2.60)	0.028 (1.61)
C	-0.0030 (-1.21)	-0.0066 (-1.01)	-0.0021 (-0.79)	-0.0046 (-0.97)	-0.0089 (-1.10)	0.0028 (0.24)	-0.0052 (-0.91)
$Experience$	-0.0011*** (-5.27)	-0.0014*** (-3.54)	-0.00070*** (-3.25)	-0.0018*** (-5.18)	-0.0010 (-1.38)	-0.0010** (-1.98)	-0.00040 (-0.90)
$Size$	0.00080 (1.32)	0.00031 (0.30)	0.00050 (0.83)	0.00070 (0.83)	-0.00030 (-0.29)	0.0015 (1.36)	-0.00050 (-0.61)
$InvestmentTiming$	-0.0067*** (-3.93)		-0.0032** (-1.99)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.10	0.12	0.12	0.087	0.14	0.16	0.15
No. of observations	2762	1225	2489	1709	910	469	929

Table 26: Excess Exit Rates and Nearby Investment Ratio

Dependent variables are fund's exit performance including excess IPO rate, excess M & A rate, and excess fail rate. Column (1) to (7) are tests on full sample, leading group, non-leading group, first round group, second round group, third round group, and fourth round and beyond group, respectively. N is fund's nearby investment ratio which follows the definition of each sample group. C is the concentration index. $Experience$ is number of fund that fund's firm has. $InvestmentTiming$ is the average of time difference between fund's first investment time of its portfolio companies and the first investment time of its portfolio companies. Unit on $Size$ is hundred million U.S. dollar in 2019 value. t-Statistics are showed in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A: Excess IPO Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N	-0.019** (-2.31)	-0.023* (-1.71)	-0.015 (-1.47)	-0.019* (-1.77)	0.024 (1.26)	-0.017 (-0.57)	-0.0083 (-0.34)
C	0.022*** (9.00)	0.012** (2.35)	0.028*** (9.10)	0.014*** (3.65)	0.019** (2.55)	0.024** (2.04)	0.029*** (3.95)
$Experience$	0.0014*** (6.59)	0.0015*** (5.01)	0.0013*** (5.35)	0.0014*** (4.78)	0.00040 (1.10)	0.0021*** (4.10)	0.0014*** (2.72)
$Size$	0.0037*** (6.41)	0.0021*** (2.70)	0.0042*** (6.30)	0.0024*** (3.38)	0.0037*** (4.07)	-0.00020 (-0.14)	0.0029*** (2.59)
$InvestmentTiming$	0.021*** (12.07)		0.015*** (8.13)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.17	0.084	0.13	0.070	0.072	0.074	0.067
No. of observations	2762	1225	2489	1709	910	469	929
Panel B: Excess M & A Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N	0.011 (0.87)	0.030 (1.39)	0.0050 (0.36)	0.030* (1.78)	-0.024 (-0.88)	-0.055 (-1.34)	0.019 (0.67)
C	0.0023 (0.16)	0.025*** (3.13)	-0.0054 (-1.28)	0.017*** (2.84)	-0.0037 (-0.35)	-0.00090 (-0.052)	-0.00080 (-0.088)
$Experience$	0.00048 (1.23)	0.00020 (0.47)	-0.00040 (-1.22)	0.0010** (2.19)	-0.000051 (-0.091)	-0.00010 (-0.18)	-0.00050 (-0.83)
$Size$	0.0022** (2.53)	0.0012 (0.94)	0.0015 (1.58)	0.0011 (1.01)	0.00010 (0.11)	0.00060 (0.38)	-0.00030 (-0.25)
$InvestmentTiming$	0.013*** (4.97)		0.012*** (4.42)				
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.058	0.064	0.055	0.051	0.035	0.079	0.073
No. of observations	2762	1225	2489	1709	910	469	929
Panel C: Excess Fail Rate							
OLS	(1)	(2)	(3)	(4)	(5)	(6)	(7)
N	0.020** (2.39)	0.043** (2.42)	0.0089 (1.03)	0.029** (2.20)	0.0044 (0.21)	0.071** (2.48)	-0.020 (-1.05)
C	-0.0034 (-1.38)	-0.0071 (-1.08)	-0.0025 (-0.94)	-0.0053 (-1.11)	-0.0097 (-1.20)	0.0026 (0.22)	-0.0045 (-0.79)
$Experience$	-0.0011*** (-5.40)	-0.0014*** (-3.63)	-0.00070*** (-3.39)	-0.0018*** (-5.25)	-0.00060 (-1.47)	-0.0011** (-2.16)	-0.00030 (-0.81)
$Size$	0.00070 (1.18)	-0.00041 (-0.04)	0.00040 (0.68)	0.00060 (0.69)	-0.00020 (-0.23)	0.0014 (1.26)	-0.00040 (-0.52)
$InvestmentTiming$	-0.0068*** (-3.96)		-0.0034** (-2.07)				
Fund Location	yes	yes	yes	yes	yes	yes	yes
Vintage Decade	yes	yes	yes	yes	yes	yes	yes
R^2	0.10	0.12	0.12	0.086	0.14	0.16	0.15
No. of observations	2762	1225	2489	1709	910	469	929

BIBLIOGRAPHY

BIBLIOGRAPHY

- Christopher J. Malloy. 2005. The Geography of Equity Analysis. *The Journal of Finance*, Volume 60, NO. 2, 719–755.
- Delphine Charlet and Geraldine Damnati. 2017. SimBow at SemEval-2017 Task 3: Soft-Cosine Semantic Similarity between Questions for Community Question Answering. *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, 315–319.
- David J. Denis, Diane K. Denis and Keven Yost. 2002. Global Diversification, Industrial Diversification, and Firm Value, *The Journal of Finance*, Volume 57, NO. 5. 1951-1979.
- Grigori Sidorov¹, Alexander Gelbukh¹, Helena Gomez-Adorno¹, and David Pinto. 2014. Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model, *Computacion y Sistemas Volume 18, No. 3*, pp. 491-504.
- Han Degryse and Steven Ongena. 2005. Distance, Lending Relationships, and Competition. *The Journal of Finance*, Volume 60, NO. 1, 231–266.
- Henry Chen, Paul Gompers, Anna Kovner, Josh Lerner. 2010. Buy local? The geography of venture capital. *Journal of Urban Economics*, Volume 67, 90-102.
- Josh Lerner. 1995. Venture Capitalists and the Oversight of Private Firms. *The Journal of Finance*, Volume 50, NO. 1, 301-318.
- Marcin Kacperczyk, Clemens Sialm and Lu Zheng. 2005. On the Industry Concentration of Actively Managed Equity Mutual Funds. *The Journal of Finance*, Volume 60, NO. 4, 1983–2011.
- Mark Humphery-Jenner. 2013. Diversification in Private Equity Funds: On knowledge-sharing, risk-aversion and limited-attention. *Journal of Financial and Quantitative Analysis*, Volume. 48, No. 5, 1545-1572.
- Morten Sorensen. 2007. How Smart Is Smart Money? A Two-Sided Matching Model of Venture Capital. *The Journal of Finance*, Volume 62, NO. 6, 2725-2762.
- Nicole Choi, Mark Fedenia, Hilla Skiba, and Tatyana Sokolyk. 2017. Portfolio Concentration and Performance of Institutional Investors Worldwide. *Journal of Financial Economics*, Volume. 123, Issue 1, 189-208.
- Paul Gompers, Anna Kovner, and Josh Lerner. 2009. I Specialization and Success: Evidence from Venture Capital. *Journal of Economics & Management Strategy*, Volume 18, Number 3, 817–844.

- Paul A. Gompers, Will Gornall, Steven N. Kaplan, and Ilya A. Strebulaev. 2020. How Do Venture Capitalists Make Decisions? *Journal of Financial Economics*, Volume. 135, Issue 1, 169-190.
- Rajarishi Nahata. 2008. Venture capital reputation and investment performance. *Journal of Financial Economics*, Volume. 90, Issue 2, November 2008, 127-151.
- Rajarishi Nahata, Sonali Hazarika, and Kishore Tandon. 2014. Success in Global Venture Capital Investing: Do Institutional and Cultural Differences Matter? *Journal of Financial and Quantitative Analysis*, Volume. 49, No. 4, August 2014, 1039–1070.
- Shai Bernstein, Arthur Korteweg, and Kevin Laws. 2017. Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *The Journal of Finance*, Volume 72, NO. 2, 509-538.
- Shai Bernstein, Xavier Giroud, and Richard R. Townsend. 2016. The Impact of Venture Capital Monitoring. *The Journal of Finance*, Volume 71, NO. 4, 1591-1622.
- Stephan Hollander and Arnt Verriest. 2016. Bridging The Gap: The Design of Bank Loan Contracts And Distance. *Journal of Financial Economics*, Volume. 119, Issue 2, February 2016, 399–419.
- Steven N. Kaplan and Antoinette Schoar. 2005. Private Equity Performance: Returns, Persistence, and Capital Flows. *The Journal of Finance*, Volume 60, NO. 4, 1791-1823.
- Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhersch, and Armand Joulin. Advances in Pre-Training Distributed Word Representations. *Computer Science arXiv:1712.09405*.
- Veronika K. Pool, Noah Stoffman, and Scott E. Yonke. 2012. No Place Like Home: Familiarity in Mutual Fund Manager Portfolio Choice. *The Review of Financial Studies*, Volume 25, NO, 2563–2599.
- Xi Han. 2009. The Specialization Choices and Performance of Venture Capital Funds. *Working paper*.
- Xiaohui Gao, Jay R. Ritter, and Zhongyan Zhu. 2013. Where Have All the IPOs Gone? *Journal of Financial and Quantitative Analysis*, Volume. 48, No. 6, 1663-1692.
- Xuan Tian. 2011. The Causes and Consequences of Venture Capital Stage Financing. *Journal of Financial Economics*, Volume 101, Issue 1, 132-159.
- Yael V. Hochberg, Alexander Ljungqvist and Yang Lu. 2007. Whom You Know Matters: Venture Capital Networks and Investment Performance. *The Journal of Finance*, Volume 62, NO. 1, 251-301.

Yael V. Hochberg, and Joshua D. Rauh. 2013. Local Overweighting and Underperformance: Evidence from Limited Partner Private Equity Investments. *The Review of Financial Studies*, Volume 26, NO. 2, February 2013, 403–451