UNIVERSITEIT GENT

FACULTY OF ECONOMICS AND BUSINESS ADMINISTRATION

Essays on Behavior and Extreme Events

Martien Lamers

2014

Submitted to the Faculty of Economics and Business Administration of Ghent University, in fulfillment of the requirements for the degree of Doctor in Economics

Essays
on
Behavior and Extreme Events

Martien Lamers

Department of Financial Economics

Ghent University

A thesis submitted for the degree of

Doctor of Philosophy at Ghent University

October 2014

This book was typeset using LaTeX.

Essays on Behavior and Extreme Events

Martien Lamers

Doctoral Jury

Prof. dr. Marc De Clercq	Dean-President, Ghent University
Prof. dr. Patrick Van Kenhove	Academic Secretary, Ghent University
Prof. dr. Michael Frömmel	Promotor, Ghent University
Prof. dr. Jaap W.B. Bos	Co-Promotor, Maastricht University
Prof. dr. Michael Kötter	Frankfurt School of Finance and Management
Dr. Rachel Pownall	Maastricht University, Tilburg University
Prof. dr. Koen Schoors	Ghent University
Prof. dr. Rudi Vander Vennet	Ghent University

If you can meet with Triumph and Disaster

And treat those two impostors just the same

Rudyard Kipling, *If–*

For my parents and Victoria

Acknowledgments

When I started this project back in 2009, I planned – partly driven by outside forces – to finish my dissertation in 2 to 3 years. Looking back 2 years after my initial deadline, I'm glad that I used the extra time at my disposal to finish and become a more complete researcher. Moreover, I've been lucky to have many great experiences during these 5 years that have helped me grow in many different ways. Obtaining a Ph.D. is a rewarding, but challenging, accomplishment. While at times it might feel like a paid vacation, others times it can also be very confrontational. You learn as much about yourself as you do about research. It is a continuous process in which you build up knowledge and ideas, tear them down and build them up again. As such, it can feel like a marathon, with ups and downs – sometimes monthly, often weekly or even daily – and you need to learn how to deal with these growing pains. A Ph.D. is therefore above all a degree in persistence and determination. Luckily, I was extremely fortunate to have great support, without whom this dissertation would have never been written. I would like to take the opportunity to thank them here.

First, I want to thank my promotor and co-author, Michael, for giving me a shot in the 'majors' after finishing the Master of Banking and Finance. Michael offered to give me one year of funding, during which we managed to obtain funding that has financed the past four years of research. You always allowed me to find my own way, and you never once denied me from attending conferences, summer schools and other courses. Thank you for that, and thank you for taking me on as a Ph.D. student.

I also want to thank my co-promotor, Jaap. Six years ago, Jaap was co-promotor of my master thesis in Utrecht. At the time I wrote that "*I can honestly say that without him this thesis would not have been completed*". These words still ring true today. For the second time in the last three theses I have written, I can honestly say that without Jaap this dissertation would not have been completed. From the moment I decided to pursue a Ph.D., Jaap helped me find a place where I could do so and gave advice during a lunch in Haarlem, with the usual 'uitsmijter' and glass of milk. As co-promotor, you have given me so many insights, not just on how to write papers, but also how to look at economic questions and frame them in the right way. You were always willing to give advice, picked me up when I needed to, helped me with coding, read my papers, asked critical questions and pushed me to become better. But there was also time for fun, the RES meetings, recently the FMA in Maastricht, as well as the many other research visit to Maastricht come to mind. I can't thank you enough for all of this, for keeping me grounded and for helping me with my job search. Besides co-promotor and mentor, you have become a great friend, and I appreciate everything you have done for me and Viki. I look forward to continuing our joint research in the future.

The last four years were funded by the Special Research Fund (Bijzonder Onderzoeksfonds) of Ghent University. I want to express my gratitude to them and to the Hercules Foundation, which funded the research infrastructure that made this work possible.

I am deeply indebted to the external members of the reading committee, Rachel Pownall and Michael Kötter. Thank you for the stimulating discussions and your comments, they have added a dimension to the chapters which they otherwise would have lacked. I am also grateful for your words of encouragement and help, and to remind me why academics do what they do. I specifically want to thank Rachel for organizing the FMA European Meetings in Maastricht this year where I picked up many tips on how to survive in academia. I made many new contacts by attending the Doctoral Consortium, and will always remember the amazing dinner in Château Neercanne. The internal members of the reading committee, Koen Schoors and Rudi Vander Vennet, were my professors in the Master of Banking and Finance. Even though they became colleagues, they never stopped teaching me and provided inspiring examples I have learned from over the years. Rudi, thank you for everything you have done for me, from reading my papers, to helping me with research proposals and giving me advice on the next steps in my career. Also, thank you for introducing me to the finer restaurants in Ghent, as well as the good bars and special beers. I am thankful to Koen for all the discussions that we've had about research and my papers. You are one of the fastest thinkers I know and you have inspired me more than you know. In seminars, you are never the first one to speak or ask questions, but when you do, you're always right on the money. What I especially admire about you is that you do not ask difficult questions without providing the answer. I will not forget the sunny terrace in Paris, and I look forward to working with you on our common project. Rudyard Kipling must have had both Rudi and Koen in mind when he wrote *"If you can think—and not make thoughts your aim"* in his poem *If–*; it describes the both of them perfectly. Thank you for everything you have done and I hope to be able to repay you one day.

I have also been lucky to do research with several co-authors. To my co-authors on the first two chapters, Valia and Adi, thank you for giving me the opportunity to join your research project, for teaching me to always dig deeper and to not be satisfied with an easy answer. I can't believe that we managed to write a couple of papers based on what started out as a simple master thesis project in Utrecht, what feels like so many years ago. Moreover, I thank Frank and Ewald for making data available and providing practical input, and Mark for inviting me to lecture in Utrecht and letting me join the project he and Jaap had in mind.

I want to thank my colleagues at the Department of Financial Economics for 'putting' up with me for so many years, as well as my new colleagues at the Department of Economics, Econometrics and Finance of the University of Groningen for giving me a warm welcome. I especially want to thank Ruud and Harry for giving me the opportunity to continue to do research in a stimulating environment.

Having a loud Dutch colleague must have been an 'interesting' experience for my colleagues. Nevertheless, you welcomed me and accepted my Dutch ways. I want to thank Nathalie and Sabine, for helping with all the administrative work that goes into research and teaching. You made the whole experience enjoyable, were always ready for a 'klapke', and taught me how to pronounce Flemish names and words. All colleagues, past and present, at FinEco created a nice work environment, which was intellectually stimulating and supportive of research. Experiencing the finishing stages together with Selien was exactly what we both needed. We picked each other up, urged each other to keep on going through difficult times, and helped to keep the administration in order. I specifically want to thank Dries, Joris, Julio, Frederick, Garo, Glenn, Valerie and Xing for always being there to talk (or complain) about research, the exploits of our favorite sportsmen/teams, for reading my papers or for just enjoying a hamburger.

I also want to thank colleagues from the departments of general economics and social economics. Matthijs, thanks for your support as my fellow 'Nederbelg'. Ewoud, I am glad we got to know each other in Göteborg, I will not forget our burger-exploit and our tour of the city with Selien. Bruno, thank you for welcoming Viki and me into your home and for your advice and help. Gerdie, thank you for taking the time to discuss methodological issues and asking tough questions during the internal seminars. A special thanks to all colleagues for the Friday night drinks at the Geuzenhuis and for the garden lunches once we all moved to the new offices in St. Pietersplein.

I also want to acknowledge support from the kind people at the College of Europe, where I used to work on Friday's while living in Bruges. Jacques, Eric, Jessie, Chiara, Federico, Renée and Delphine, thank you for the discussions, the lunches and the drinks at De Garre! Without the College of Europe I would not have met Christophe, who helped and advised me regarding the future, and who I know will be a valuable co-author in the future.

Despite the fact that Ghent is only a 2-hour drive from Utrecht, keeping in touch with the home front is always a bit challenging. I want to thank the people that kept in touch, and apologize for being unavailable the last couple of months/years. Talking to people outside of the protective shell of academia was necessary to keep my sanity. To my old study-friends Florian, Harry, Coos, Rienk and Onno, thanks for coming by and reviving the old Utrecht-spirit. Thanks to my old and my new colleague Lu, I'm looking forward to catch up with you in Groningen. Thanks to Abe and Marco, Viki and I are so lucky to call you our friends. Even though they are not technically on the home front, I want to thank Gary and Jeff. I've known Gary since I was a little boy, and he is the person who instilled in me the love for baseball in general and specifically the Yankees. To him and his son, Dr. Jeff, thanks for the fun times, I very much enjoyed New York and the time we met Jeff in Amsterdam.

A small acknowledgment is also in order for Jorge Cham, creator of the best web comic depicting the life of graduate students, `www.phdcomics.com`. Knowing that others are dealing with the same problems, even though they are fictional characters, can help you to keep on fighting, although it also provided a very dangerous procrastination tool!

My family has always stimulated and encouraged me to continue my education, and they also created the environment that made it possible to pursue this goal. My sister Monica and her longtime partner Aaldert, the brother that I never had, thank you for all your help. Viki and I are so thrilled that Mathias has joined our little family. Thank you for helping us move so many times, for helping us through bad times and for just being there. My father is the reason that I studied economics. I still remember the first time I 'interviewed' him for an assignment in school, and he explained to me that economics is a science determining how scarce resources are best used. Thank you for introducing me to economics, finance and banking through the books of Michael Lewis and Roger Lowenstein, which greatly contributed to me following this path. Both my father and mother have supported me all the way throughout university, helping me when I invariably moved around, and providing a great environment for me to continue to improve myself. Thank you for all of this, and for helping me find my motivation when I lost it.

Lastly, I want to thank my wife Viki. Simply put, this dissertation could not have been written without you. You've been here for the good, and you've helped me through the bad by believing in me and by simply being yourself. The last few months have been quite hectic, but you have smoothed them without making it look like even a small effort. You created the best environment for me to excel in, and therefore this thesis is at least 50% yours (without taking into consideration that we married in community of property). Besides reading my work and giving me comments, we co-authored one of the chapters and I hope we can do it again many more times in the future. I hope to be able to repay you for the effort you put into this dissertation. I'm glad that this opportunity will come soon, since you will be going through this in the next year. And I know that you will come through with flying colors.

Martien Lamers
Groningen, September 2014

Table of Contents

List of Figures

List of Tables

1 Introduction

Many standard economic models that are used to study the behavior of (groups of) economics agents assume that these agents have rational expectations. For instance, the *Efficient Market Hypothesis* – developed in seminal papers by Samuelson (1965), Fama (1965, 1970) and later extended with micro-foundations by e.g. Lucas (1978) – assumes that the population of economic actors is on average correct and updates expectations (and thus prices) properly when confronted with new information. This implies that investors have intimate knowledge of the underlying probability distribution of relevant events as well as their effect on returns. However, a long line of research in the field of behavioral finance has criticized these assumptions by documenting several behavioral biases people suffer from when making financial decisions. For excellent surveys see, e.g., Malkiel (2003) and Lo (2007). Perhaps the most prevalent biases are documented by Tversky and Kahneman (1973) and Kahneman and Tversky (1979), stating that people do not properly evaluate small probability events and that events that are more easily remembered, or salient, are judged to be more common. Notwithstanding these biases, the Efficient Market Hypothesis does not assume that individual investors all have correct expectations, rather that the entire population – the market – is correct. The question I look at in this dissertation is to what extent these individual biases influence aggregate behavior. Does the population of agents also have these biases, i.e. are the individual biases aggregated into a group bias, or do biases average out in a group and therefore disappear?

Experimental findings, by e.g. Bottom et al. (2002), have shown that most individual biases tend to disappear in a group setting, implying that assumptions on which the Efficient Market Hypothesis relies are correct. However, the authors find that biases related to extreme events are actually propagated into group behavior, meaning that biases documented by Tversky and Kahneman (1973) and Kahneman and Tversky (1979) do not disappear. When confronted with extreme events such as terrorism or natural disasters – both of which have a low probability, a high but uncertain impact, as well as a salient nature – the stage is thus set for overreaction by economic actors (see e.g. De Bondt and Thaler, 1985, 1990).

This dissertation focuses on the behavioral effects in financial decision making of extreme events, leaving the controlled setting of a laboratory in order to identify whether individual biases relating to extreme events can be found in aggregate (price) data. To do so, I focus my research in three lines and ask the following broad questions: How do economics agents take extreme events into account ex-ante when making decisions? How do extreme events change aggregate behavior of agents ex-post? And how can we minimize the impact of extreme events?

The view that the world is an inherently uncertain and probabilistic place has been popularized in recent years, raising awareness that that we only see one realization of the underlying data-generating process (see e.g. Taleb, 2007). Extreme events are those events that fall in the tail of the distribution: although there is a small probability of occurring, they have a potentially large (economic) impact. The September 11 attacks are an example of such an extreme event, killing 3,000 people and causing damage estimated to be well over a trillion dollars (see IAGS). The unlikely nature and magnitude of these attacks become apparent when considering that, aside from the 9/11 attacks (GTD), there have been only 1,000 casualties related to terrorism in the United States since 1990. As noted, extreme events come to mind quickly and are therefore likely to be judged more common than they are in reality, with all kinds of consequences. A classic example of irrational behavior of people with respect to small percentage events is given by Johnson et al. (1993), who found that airplane passengers were willing to pay more for life-insurance covering them against terrorism then for insurance covering any cause. Interestingly, even though agents have a short memory, research has suggested that not only do investors act on (salient) extreme events when they occur (Klibanoff et al., 1998; Hirshleifer and Teoh, 2003), they are also likely to remember the event and act accordingly throughout their lifetime (Malmendier and Nagel, 2011).

Hence, extreme events have the power to change behavior in the short- *and* in the long-run. However, events such as terrorism, natural disasters or financial crises are not that infrequent at all: over 4,000 natural disasters have taken place worldwide since 1980 (EM-DAT) and around 16,000 incidents of terrorism have been recorded between 1998 and 2010 alone (GTD). Perhaps even more surprisingly, Laeven and Valencia (2012) document roughly 150 banking crises and 200 currency crises worldwide since 1970. Yet when asked to consider these types of events, most people tend to think about the attacks on 9/11, London or Madrid; the recent financial crisis or the Great Depression; the Indian Ocean tsunami, hurricanes Katrina, Sandy or more recently typhoon Haiyan. While they may not have the magnitude of these larger extreme events, the smaller events are plentiful and, analogous to the 'anchoring effect' introduced by Tversky and Kahneman (1974), can still lead to behavioral biases by reminding people of the larger, salient events.

This dissertation positions itself against this background and looks at how economic actors take extreme events into account ex-ante, how it changes their behavior ex-post and how the impact of extreme events can be minimized. In the remainder of this introductory chapter, I outline the studies undertaken and briefly summarize their results.

1.1 Pricing decisions, forecasting and extreme events

The first two chapters focus on how economic actors take extreme events into account ex-ante, and investigate the behavior of a market maker who is faced with two sets of counter-parties in the population: noise traders and insider traders. Since the last group has superior knowledge compared to the market maker, the latter adjusts his prices such as to minimize the impact insiders can have on his profits. The market considered in these chapters is a fixed-odds betting market, which encompasses direct bets on winning or losing and a well-defined endpoint, thereby making it extremely suitable for testing hypotheses developed for financial markets (Vaughan Williams, 1999). Since this implies that the market maker (read: bookmaker) takes sizeable positions on the outcome of each bet (Levitt, 2004), she therefore has an incentive not to lose money to insiders who might be better informed. Chapter 2 develops a model of how a bookmaker will determine her odds when considering noise traders, insiders and a population consisting of both. In line with Silver (2012), I take the view that insiders use Bayesian reasoning and do not necessarily know the identity of a winner, but rather that they have a better understanding of the true winning probabilities compared to the bookmaker. This represents a break from the existing literature which assumes that insiders know the identity of the winner (see e.g. Shin, 1991, 1992, 1993). The model presented in Chapter 2 shows that the odds quoted by the bookmaker are equivalent to offering a (real) call option to the insiders: the higher the expected mispricing, the higher the chance that insiders will exercise this option. This is especially important for bets concerning longshots (with a low chance of winning but a high possible payout), as the bookmaker is faced with possible large losses compared to bets on favorites (with a high chance of winning but a low payout). The model predicts that due to the presence of insiders, bets on longshots are made less attractive in order to avoid this high payout. In this sense, the model is consistent with, and can explain the existence of the favorite-longshot bias in betting markets (see e.g. Vaughan Williams and Paton, 1997; Snowberg and Wolfers, 2010), which states that odds are biased estimates of winning probabilities as favorites (longshots) are consistently given a lower (higher) probability of winning than in reality occurs. By making assumptions on the data-generating process of news inflows, I simulate the options values that the bookmaker offers the pool of bettors. I calibrate the model on quoted odds for the 1998 Australian Horse Racing season and find that the bookmaker expects 97 percent of bets to be places by insiders. However, when combined with more objective measures on the extent of insider trading, I find that only roughly 60% of the bets consisted of insiders and 40% of noise traders. The bookmaker thus overestimates to a large extent the amount of insiders and the possible losses to them, which is consistent with the inability to process small probability events as documented by Tversky and Kahneman (1973). In Chapter 3 I find that while the generated option values do increase the forecasting power beyond the odds that are offered by bookmakers, which incorporate all publicly available information, a trading rule based on them was found to have negative proceeds. While the option values do contain information that is not reflected in the prices set by the bookmaker, this information is not economically significant, effectively proving that the market under investigation is weak-form efficient.

1.2 Investor behavior and extreme events

Having seen how economic actors take into account extreme events ex-ante when making financial decisions, Chapter 4 focuses on how extreme events can change their behavior ex-post. Previous research has shown the tendency of extreme events to have a short-run, as well as a longlasting impact on prices and behavior. One extreme event that has been researched extensively is the attack on September 11, 2001 and its impact on, amongst others, financial markets.[1] It has been found that stock prices react to extreme events such as terrorism (see, amongst others, Chen and Siems, 2004; Drakos, 2004; Eldor and Melnick, 2004; Karolyi and Martell, 2010; Chesney et al., 2011), but these reactions show heterogeneity across attacks, sectors of the economy and countries. More specifically, although it has been established that terrorist attacks can lead to reactions in the countries they occur in, it is unclear how and through which channel they affect stock markets in other countries. Recently Abadie and Gardeazabal (2003) and Drakos (2010a) have suggested that economic linkages between countries/regions might be the channel through which the shocks of terrorist attacks are transmitted to other countries/regions. However, the latter uses events in a post-9/11 world, disregarding the possibility that the relationship might have changed over the years. In the chapter, I find evidence consistent with the channel of economic linkages: U.S. equity prices drop after large terrorist attacks abroad, and the drop is proportional to the capital stock of U.S firms. When estimating the relationship on subsamples, I find that the relationship is only significant, both statistically and economically, *after* the tragic events of September 11, 2001. Multiple explanations are explored, but the most likely explanation is a disaster myopia on behalf of U.S. investors prior to the attacks, followed by a pronounced increase in sensitivity to attacks on foreign soil after. Inter-

[1] For an extensive survey on the impact of terrorism and 9/11 on financial markets and many other aspects see Frey and Luechinger (2005).

estingly, the reactions after the attacks are rational since they are proportional to the stock of investment by U.S. companies. This suggests that investors do not overreact but have rather experienced a wake-up effect following the attacks, resulting in more awareness of the dangers of terrorism on foreign soil for U.S. multinationals.

1.3 Depositor behavior, bank size and extreme events

The final two chapters of the dissertation are inspired by the extreme events that occurred in the United States during 2007-2008. The financial crisis forced many large financial institutions into bankruptcy, as it did with Lehman Brothers, and even threatened to collapse the entire financial system. While the general public watched on as governments around the world had to bailout these institutions, smaller banks were allowed to fail. Approximately 400 commercial banks were closed down by the Federal Deposit Insurance Corporation (FDIC) since the beginning of the crisis and another 500 banks in distress were forced to merge with healthier counterparts. Chapter 5 looks at depositors, who are trusted by regulators to monitor their banks and move their deposits to other banks if they feel that their current one is too risky. Both government interventions and bank failures can give mixed signals to depositors regarding the need to monitor banks. On the one hand, bailouts have weakened incentives to monitor as depositors trust the government to intervene when their deposits are at stake. On the other hand, previous literature has documented that experiencing events such as banks failures can lead to an increase of risk aversion among depositors (see e.g. Martinez Peria and Schmukler, 2001; Karas et al., 2010, 2013; Iyer and Puri, 2012). After these events, depositors generally increase their monitoring of banks. Given that both types of events occurred, it is an empirical question which effect dominates. In Chapter 5, I find that in regions that suffered from bank failures, depositors monitor their banks more intensely, despite the existence of bank bailouts. Moreover, in regions with more bank failures, there is more discipline altogether.

Finally, Chapter 6 tries to draw a lesson from the crisis in order to prevent a new one. Since many institutions had to be bailed out, thereby burdening public finances, much effort has been focused to avoid another crisis of this kind. The emphasis in this discussion is often placed on the interconnectedness, its level of diversification, or the size of a bank (see e.g. De Jonghe, 2010; Zhou, 2010; Adrian and Brunnermeier, 2011; Markose et al., 2012; Bertay et al., 2013). Taking the view that the supervisor is a portfolio manager who holds a portfolio of banks, this chapter asks what the optimal portfolio structure should be in order to reduce the portfolio (systemic) risk and therefore the likelihood of a new crisis. I find that the levels of concentration in the current portfolio are consistently too high and that a less concentrated banking system could see lower levels of systemic risk.

1.4 Suggestions for further research

These chapters are united by the same biases in behavior as described in Tversky and Kahneman (1973) and Kahneman and Tversky (1979), namely improperly evaluating small probability (or extreme) events and judging them to be too uncommon. This dissertation shows that these biases can be found in aggregate financial decision making across many different types of events, and that they can persist for a long time. As such, it provides evidence that standard models relying on rational expectations of a group of agents suffer from the same biases relating to extreme events that individual suffer from. Future avenues of research need to take these findings into account and focus on understanding the group dynamics that occur. Moreover, future research is also needed to understand if other biases, that were found to average out by experimental literature (see e.g. Bottom et al., 2002), can also be found in aggregate price data.

2

Pricing Decisions and Insider Trading in Horse Betting Markets*

2.1 Introduction

This chapter looks at the decision a bookmaker makes when setting prices under uncertainty in a fixed odds betting market. We argue that this decision can be modeled in a call option framework to measure the degree of insider trading in racetrack betting markets. Makropoulou and Markellos (2011) first developed an option-pricing framework for the pricing of bets in fixed-odds markets and in particular for the European soccer betting market. In this market, the odds are offered by bookmakers via fixed-odds coupons several days before the game and they remain largely unchanged throughout the betting period. Their model deals with expert traders who either exploit public information in a manner superior to that of bookmakers or obtain access to new public information sooner than bookmakers do. Our approach differs in that we focus on racetrack betting where odds change frequently during the half an hour betting period. In our context public information is irrelevant since it can be incor-

*This chapter is based on Schnytzer, Lamers & Makropoulou (2010a) and Schnytzer, Makropoulou, and Lamers (2014), published in the *Journal of Gambling Business and Economics* and the *Oxford Handbook on the Economics of Gambling* respectively.

porated into new odds as soon as it hits the market. On the contrary, we deal with insiders who possess private information. Of course, the implications of trading with insiders in the racetrack betting market where the bookmaker changes his odds frequently can be quite similar to those of trading with experts in a market where the odds remain unchanged. However, the MM framework could not be readily applied to the racetrack betting market due to structural differences between the two markets. In order to fill in this gap, Schnytzer, Lamers & Makropoulou (2010a) developed a model for the pricing of bets in a market with insiders relying on the MM framework and applied it to the Australian racetrack betting market. In this chapter we extend the work done by Schnytzer, Lamers & Makropoulou (2010a) in several aspects. First, we relax their assumption of continuous arrival of information by employing a more realistic specification in which information arrives in discrete amounts and therefore the true probability of a horse winning exhibits quantum jumps and dives. Second, the model is extended to allow for more periods in which betting takes place. Whereas the previous work of Schnytzer, Lamers & Makropoulou (2010a) assumes a betting period in which the bookmaker sets his prices once (at opening prices), we extend the model to accommodate more time periods in which a bookmaker quotes prices. More specifically, we follow the data at our disposal and allow for betting by insiders both at opening prices and middle prices, instead of only at one of those. Finally, to derive the probability of insider trading, the zero profit condition of the bookmaker does not have to hold for every single horse. This condition is necessary only for the race, allowing the bookmaker to make losses only on the horses he expects insiders to bet on and to make profit on the horses that are backed by outsiders, that bet according to subjective winning probabilities in accordance with public information as explained below.

The way in which insiders bet involves the so-called plunge. This is where several agents of the insider approach different bookmakers simultaneously and back the same horse at the best odds available. The reason why a single bettor is usually insufficient is that bookmakers are permitted to refuse bets which would leave them with large contingent debts.[1] Accordingly, an insider wishing to place really large sums of money on a particular horse will need to spread the bet across bookmakers. It should

[1] The precise size of the maximum bet which a bookmaker must accept varies from place to place, but is rarely above a thousand dollars.

be noted that on-course bookmakers are small, independent firms who compete in selling a homogeneous product. Accordingly, competition among them is fierce and the trend in prices during the betting is always downwards unless a horse is plunged.[2] However, since all bookmakers need to determine initial odds and all bookmakers in Australia must be members of the relevant state bookmakers' Association, they tend to save on research costs by obtaining a set of opening odds from the Association. These are not obligatory, but tend to be widely used. The important thing about these prices is that they contain a high built-in expected mark-up which serves as a cushion of sorts against insiders. Of course, once a plunge arrives, every bookmaker is on his/her own, and the prices of all horses in the race fluctuate freely. For our purposes, any fall in odds (increase in price) is taken as evidence of a plunge. Similar to Schnytzer, Lamers & Makropoulou (2010a), we combine our theoretical measure of insider trading with data on plunging and find the degree of insider trading to be 60% in our dataset.

The remainder of this chapter is organized as follows. In the second section we discuss the general framework. In section 3 we will build the theoretical model and the empirical model is discussed in section 4. Finally, in section 5 we will present our findings.

2.2 General framework and model assumptions

Our objective is to build a model of bookmaker optimal pricing, assuming that there are two populations of bettors in the market; namely, outsiders and insiders. We begin by describing the general framework and main assumptions with respect to the information possessed by the market agents and their betting criteria along with the trading process and the pricing response by bookmakers.

Assume there are N horses in a race. The problem of the bookmaker is that of determining the opening odds. We denote by $\theta_j(0)$ the odds quoted by the bookmaker at time 0 against horse j winning, where $j = 1, ..., N$. If a bet is successful then, ignoring taxes, the bettor receives back $1 + \theta_j(0)$ on a one unit bet. An opening price $OP = \phi_j(0)$

[2] For the purposes of this chapter, prices are defined in their economic sense as the amount which must be bet on a horse to ensure a total payback (including the initial outlay) of $1. Odds, on the other hand, have their traditional meaning; i.e., if the odds on a winning horse are X to 1, then a $1 bet on the winning horse yields a total payback (including the $1 outlay) of $X + 1$.

implies odds $\theta_j(0) = \frac{1-\phi_j(0)}{\phi_j(0)}$.

Suppose also that the horses' true winning probabilities at any point in time t are given by $P_j(t)$, $j = 1, ..., N$ where $\sum_{j=1}^{N} P_j(t) = 1$. These true winning probabilities are assumed to evolve according to the flow of information, both public and private, throughout the betting period until the race starts and are therefore stochastic. Moreover, we assume that the flow of information is tied to the flow of bets. In this sense new information is said to have hit the market only if bets arrive in the marketplace in a way that alters the horses' winning probabilities, as those were until then perceived by the bookmaker. The stochastic process for the true winning probabilities could be either continuous or discontinuous, i.e. a jump process, or a mixture of the two. Strictly speaking, the process that affects the true probability should be seen as discontinuous, since the flow of information from small events that may affect the outcome of the race is not continuous. Moreover, we assume that the flow of public information during the betting period is negligible, at least compared to the amount of private information that may hit the market. This assumption makes sense especially if one considers the nature of racetrack betting and the short betting period (about 30 minutes). Moreover, it implies that whenever the bets arrive in a way different from the bookmaker's expectation, it is due to trading on inside information, unknown to the bookmaker until the actual trade has taken place. The above suggest that the expected value of $P_j(t)$ at any point in time, $E\left[P_j(t)\right]$, should be equal to the initial value $P_j(0)$.

Regarding the information possessed by the two presumed populations of bettors, namely outsiders and insiders, and their betting behavior, we make the following assumptions. Firstly, nobody, not even an insider, knows in advance which horse will win the race, in contrast to Shin (1991, 1992, 1993) who assumed that insiders know which horse will win the race. Secondly, an insider knows only the true winning probability of one horse k, $\widehat{P}_k$, before this knowledge becomes public. However, she does not know how $1 - \widehat{P}_k$ is distributed among the other horses. Given the quoted opening price for horse k, $\phi_k(0)$, this true winning probability might involve a profit opportunity for the insider. A risk-neutral insider will wager on horse k only if she expects a positive return. The expected return of the insider on a one unit bet is the expected value of either $\left(-1 + \frac{\widehat{P}_k}{\phi_k(0)}\right)$ or zero, whichever is greater, since the insider bets only if $-1 + \frac{\widehat{P}_k}{\phi_k(0)} > 0 \Leftrightarrow \widehat{P}_k > \phi_k(0)$. This is similar to saying that bookmakers actually offer insiders (call) options on the horses. Obviously, it is in the bookmakers' interest to offer net out-of-the money options. However, when they err by underestimating a particular horse's true winning probability, they are liable to offer a net in-the-money option on this particular horse, which the insider (who knows her horse's true winning probability) will be glad to snap up.

Outsiders have access only to public information regarding past performance and current conditions. Therefore, we would expect outsiders to support the horses in proportion to the winning probabilities implied by "public information", $P_j(0)$, which are equal to the expected values of the true winning probabilities at the closing of betting, $E\left[P_j(T)\right]$. However, in reality the winning probabilities perceived by the outsiders should also account for their attitudes towards risk as well as for the existence of any behavioral biases among them. Consequently, outsiders are assumed to support the horses in proportion to their subjective winning probabilities, denoted by $\pi_j(t)$. A favorite-longshot bias may arise if bettors are risk-loving (e.g. Quandt, 1986) or due to behavioral biases such as those considered by Kahneman and Tversky (1979). There may of course also be herding which would lead to plunge horses being overbet. The bookmakers are also assumed to know the horses' winning probabilities implied by "public information", i.e. $E\left[P_j(t)\right]$. Compared to outsiders, bookmakers are particularly skillful in gathering and processing public information and are therefore assumed to also know the marginal density function of each horse.[3] In addition, we assume that the bookmaker can accurately predict the expectations of outsiders, i.e., the outsiders' subjective probabilities are known with certainty to the bookmaker.

Trading proceeds in a number of stages. At time zero the bookmaker declares the opening prices (OP), $\phi_j(0)$, based on his perception of the true winning probabilities at this time, $P_j(0)$. At this first stage, a proportion of the outsiders bet in the market at the OP set by the bookmaker. Suppose now that a private signal revealed to a group of insiders indicates that the true winning probability of k is actually higher than the quoted OP, i.e. $\widehat{P}_k > \phi_k(0)$. The insiders will then bet on this horse, say at time t^*. Note that such signals indicating mispricing might be revealed for more than one horse. The bookmaker observes the insider betting pattern and therefore the new

[3] As we will see in Section 4, knowing the marginal density function is equivalent to knowing the volatility of the jump size and the Poisson arrival rate.

value of the true winning probability and adjusts his prices accordingly. At the other stages, the rest of the outsiders bet at the new updated prices. Note that insiders are faced with a timing dilemma. To understand this, suppose that there are two such groups of insiders, each wishing to plunge their own horse. Since a plunge reduces the prices of other horses, each group has an incentive to wait for the other to plunge first. Insiders must utilize any special information they have during the betting, since it loses all value once the race starts. Furthermore, since insider trading is both legal - only jockeys are forbidden to bet - and takes place at fixed prices, insiders have no incentive to hide their trading behavior from outsiders. Moreover, since the insider information concerning any given horse is likely known to more than one person, the longer insiders wait, the greater the risk that the information will leak to a third party. The recipient of the leak will then plunge the horse and the group of insiders - except perhaps the one responsible for the leak - may be left with odds at which betting is no longer worthwhile (see also Schnytzer and Shilony, 2002).

In the option pricing framework developed in this chapter to model the effect of information asymmetries on prices, we do not account for competitive interactions among insiders since this would increase significantly the complexity of the problem in hand while offering limited additional insight. For simplicity, we assume that insiders will place their bet once they receive the private signal.[4]

Price updating effectively continues until the last stage at which starting prices (SP) are determined as the equilibrium prices observed in the market at the end of betting. Since in contrast to the British market there is no legal SP betting in the Australian market, these prices may be assumed to embody all the available useful information regarding the race's outcome. Although price updating might actually take place several times throughout the betting period, our empirical analysis considers only three stages, the first, an intermediate and the last stage, at which opening prices (OP), middle prices (MP) and starting prices (SP) are set, respectively.

The chapter develops a model of bookmaker pricing that can be used to derive not only the OP but also any intermediate prices. At each point in time, the prices are

[4]One way to capture potential value erosion of the option due to other insiders would be to incorporate a dividend yield. According to the theory of options, it is never optimal to exercise an American option before maturity in the absence of dividends. This means that, in our context, insiders would always bet at the last minute. It is the presence of other insiders (dividends) that makes it optimal to bet before maturity.

modeled as the equilibrium of a perfectly competitive bookmaker market. Specifically, the bookmaker is assumed to be risk-neutral, (i.e. an expected profit maximizer) and there is free entry in the market. Thus, the long-run competitive equilibrium will be established when all bookmakers earn zero expected profits in the market corresponding to each race. Moreover, assuming perfect competition allows for the simplifying assumption of inelastic outsider demand. Note that if the bookmaker were a monopolist and demand were totally inelastic, maximizing profits would lead to unbounded prices.[5]

Insiders are assumed to have a collective wealth W_i. When bookmakers price horses according to the methodology developed in this chapter, they assume that insiders bet to the full extent of their wealth W_i should the opportunity arise and that W_i is evenly distributed among insider horses. Therefore, in a race of N horses, up to $\frac{1}{N}W_i$ can be placed by insiders on each horse.

We do not make any assumptions concerning the likelihood of inside traders *vis-à-vis* either favorites or longshots. Finally, transaction costs are assumed to be negligible.

2.3 The theoretical model

2.3.1 Development of the mathematical model

The problem of the bookmaker is that of determining the opening odds $\left(1+\theta_j(0)\right)$ for each one of the N horses in a race, such that his expected profit is equal to zero. Assume for the moment that only outsiders exist in the market. Then, ignoring the time-value of money, the expected profit of the bookmaker at time zero (stage 1) is equal to the total amount of money, W_n, bet by outsiders at stage 1 on the N horses minus the amount of money that the bookmaker is expected to pay out to the winners. Assume also that $w_{n,j}$ is the amount bet on horse j, where $j = 1, 2, ..., N$ and $E_0\left[P_j(T)\right]$ is the expected value of the true winning probability of horse j at the end of the betting period (time T). Note that as explained in the previous section, the proportionate amount of money bet by outsiders on each horse, $\frac{w_{n,j}}{W_n}$, is known to the bookmaker. Regarding the true winning probabilities, these might change throughout the betting period since they evolve according to the flow of information, public and private, as

[5] A formal proof can be found in the appendix.

this information is revealed through the flow of bets. However, in the absence of insiders and under the assumption that the flow of public information during the betting period is negligible (see section above), then $E_0\left[P_j(T)\right] = P_j(0)$. The expected profit of the bookmaker is:

$$E_0(\Pi) = W_n - \sum_{j=1}^{N} P_j(0) w_{n,j} \left(1 + \theta_j(0)\right) \tag{2.1}$$

Setting $\phi_j(0) = \frac{1}{1+\theta_j(0)}$, where the notation $\phi_j(0)$ is used to denote opening prices (OP), we obtain:

$$E_0(\Pi) = W_n - \sum_{j=1}^{N} \frac{P_j(0)}{\phi_j(0)} w_{n,j} \tag{2.2}$$

Given that $\sum_{j=1}^{N} P_j(0) = 1$ then, for the bookmaker to have a zero expected profit, it is sufficient that for each j the OP satisfy the following equation:

$$\phi_j(0) = \frac{w_{n,j}}{W_n} \tag{2.3}$$

Therefore, if only outsiders exist in the market and, as assumed earlier, the bookmaker can accurately predict their expectations then, for the latter to have zero expected profit on each horse, it is sufficient that opening prices are set equal to the expectation of the bookmaker about the proportion of money bet on each horse i.e., $\phi_j(0) = \pi_j$, where $\pi_j = \frac{w_{n,j}}{W_n}$ is the winning probability of horse j as perceived by outsiders. Considering that π_j actually reflects outsiders' beliefs as those are shaped by public information, risk attitudes and behavioral biases then, under the assumption that the flow of public information is small, there is no reason for the opening prices to change during the betting period.

Suppose now that insiders also exist in the market. Obviously, the final distribution of bets will depend upon both the expectations of outsiders and insiders. The bookmaker can predict with accuracy the expectations of outsiders but not those of insiders, since the latter are shaped according to the private information they receive; moreover, this information is revealed to the bookmaker only after an inside trade has taken place.

Assume again that the bookmaker gives at time zero (opening) prices $\phi_j(0)$ for each one of the horses and that the betting period is again T periods of time. It is assumed that only part of outsiders will bet at OP, $\omega_n^{OP} = \frac{W_n^{OP}}{W_n}$, while the other part will bet at later stages after they have observed insider behavior. A risk-neutral insider will wager on horse k only if she expects a positive return. The expected return of the insider on a one unit bet is the expected value of either $\left(-1 + \frac{\widehat{P}_k}{\phi_k(0)}\right)$ or zero, whichever is greater, since the insider bets only if $-1 + \frac{\widehat{P}_k}{\phi_k(0)} > 0 \Leftrightarrow \widehat{P}_k > \phi_k(0)$.

Under the above assumptions, the bookmaker is always expected to lose from trading with insiders. In particular, the bookmaker's expected loss to an insider at time zero on a one unit bet (placed at time t^*) is:

$$E_0(\Pi) = -E_0\left[\max\left(-1 + \frac{\widehat{P}_k}{\phi_k(0)}\right), 0\right] \tag{2.4}$$

It holds that $\widehat{P}_k = P_k(t^*) \neq P_k(0)$, where $P_k(t^*)$ is the true winning probability of horse k at the time the insiders place their bet (which is now revealed to the bookmaker).

The expected profit of the bookmaker is therefore:

$$E_0(\Pi) = W_n^{OP} - \sum_{j=1}^{N} \frac{E_0\left[P_j(T)\right]}{\phi_j(0)} w_{n,j}^{OP} - \sum_{j=1}^{N} w_{i,j}^{OP} E_0\left\{\max\left(-1 + \frac{P_j(t^*)}{\phi_j(0)}\right), 0\right\} \tag{2.5}$$

where $w_{i,j}^{OP}$ is the amount of money bet by insiders at OP on horse j. Note that now that insiders also exist in the market, the bookmaker cannot know what the true wining probability will be at the end of the betting period. However, as explained in the previous section, the bookmaker is assumed to know the expected value of the true winning probability $E\left[P_j(t)\right]$ at any time t.

The expression above is complicated by the fact that t^* cannot be known *a priori* to the bookmaker and hence it should be treated as stochastic. In order to simplify this, we assume that private information that may alter the true winning probability of a given horse may arrive only once for each horse. Then, we can safely state that $\widehat{P}_k = P_k(t^*) = P_k(T)$, where $P_k(T)$ is the value of the true winning probability at the closing of betting since as we said earlier private information regarding a certain horse may arrive in the marketplace only once. Of course, one might argue that the true winning probability of horse k may be lowered if at a later time new (positive) information regarding a second horse r hits the market, implying, $\widehat{P}_r > P_r(0)$. Obviously, this would always be true in a race of two horses only. However, in a race of many horses, one could accept the supposition that this new information would reduce the true winning probabilities of all other horses (for which no inside information has hit the market) except for horse k.

Given that $w_{i,j}^{OP} = \frac{1}{N} W_i$, for the bookmaker to have zero expected profit, the following condition must be met:

$$1 - \sum_{j=1}^{N} \frac{w_{n,j}^{OP}}{W_n^{OP}} \frac{E_0\left[P_j(t)\right]}{\phi_j(0)} = \frac{1}{N} \frac{W_i}{W_n^{OP}} \sum_{j=1}^{N} E_0 \left\{ \max\left(-1 + \frac{P_j(T)}{\phi_j(0)} \right), 0 \right\} \tag{2.6}$$

or

$$1 = \sum_{j=1}^{N} E_0\left[P_j(T)\right] \left\{ \frac{w_{n,j}^{OP}}{W_n^{OP}} \frac{1}{\phi_j(0)} + \frac{1}{N} \frac{W_i}{W_n^{OP}} \frac{E_0\left\{\max\left(-1 + \frac{P_j(T)}{\phi_j(0)}\right), 0\right\}}{E_0\left[P_j(T)\right]} \right\} \tag{2.7}$$

Given that $\sum_{j=1}^{N} E_0\left[P_j(T)\right] = 1$, for the above equation to hold, it is sufficient that the opening price of each horse j satisfies the following equation:

$$\frac{w_{n,j}^{OP}}{W_n^{OP}} \frac{1}{\phi_j(0)} + \frac{1}{N} \frac{W_i}{W_n^{OP}} \frac{E_0\left\{\max\left(-1 + \frac{P_j(T)}{\phi_j(0)}\right), 0\right\}}{E_0\left[P_j(T)\right]} = 1 \tag{2.8}$$

or multiplying with the term $\left(\frac{W_n^{OP}}{W_n}\right)$ and rearranging we obtain:

$$\left(\frac{W_n^{OP}}{W_n}\right) - \left(\frac{W_n^{OP}}{W_n}\right) \frac{w_{n,j}^{OP}}{W_n^{OP}} \frac{1}{\phi_j(0)} = \frac{1}{N} \left(\frac{W_i}{W_n}\right) \frac{E_0\left\{\max\left(-1 + \frac{P_j(T)}{\phi_j(0)}\right), 0\right\}}{E_0\left[P_j(T)\right]} \tag{2.9}$$

The left hand side of the above equation is the expected bookmaker gain from trading with outsiders while the right-hand side is his expected loss to insiders. The optimal price is the one that equalizes the gain from outsiders to the loss to insiders. It can be found by solving the above equation through trial and error, given the proportion of outsiders who bet at OP, $\omega_j^{OP} = \frac{W_n^{OP}}{W_n}$, outsider's subjective probabilities, $\pi_j^{OP} = \frac{w_{n,j}^{OP}}{W_n^{OP}}$, the bookmaker's expectation about the true winning probability at the closing of betting, $E_0\left[P_j(T)\right]$, the number of runners in a race, N, and, of course, the degree of insider trading (as perceived by the bookmaker) defined as the ratio of total insider money to total outsider money, $\frac{W_i}{W_n}$.

Note that the left-hand side of this equation should be greater or equal to zero since the right-hand side is always non-negative. Therefore, if insiders exist in the market then, in order for the bookmaker to have zero expected profit, prices should be set greater than outsiders' subjective probabilities, i.e.:

$$\phi_j(0) \geq \frac{w_{n,j}^{OP}}{W_n^{OP}} = \pi_j^{OP} \tag{2.10}$$

To summarize, our model suggests that since any private information is conveyed to the bookmaker only after an informed trade takes place, the latter should include a premium in the OP to compensate him for this risk. Moreover, this premium is related to the cost of trading with insiders, which in turn is a function of the degree of insider trading ($\frac{W_i}{W_n}$) and the potential value of private information that may be exploited by insiders (as captured by the term $E\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(0)}\right),0\right\}$). Under these considerations, the sum of OP would always be greater than one.

Suppose now that at a later point in time, time τ (stage 2), after the bookmaker has observed insider trading, he will set new prices (called MP). For those horses on which insider trading has taken place, say m horses, prices will be set equal to the horses' new true winning probabilities (in the absence of any bookmaker margin). The reason is that insiders pose no further risk to the bookmaker since they can only bet at either OP or MP on a given horse but not both.[6] For the rest of the horses ($N-m$), the bookmaker will set prices as above. Therefore, at the second stage the total amount of money available by insiders is $W_i - \frac{m}{N}W_i \leq W_i$. Thus, the bookmaker will quote MP as if $\frac{1}{N-m}\left(W_i - \frac{m}{N}W_i\right) = \frac{1}{N}W_i$ would be wagered by insiders on each one of the remaining $N-m$ horses should the opportunity arise. Therefore, we have:

[6]If they bet at OP then prices will exhibit a plunge and therefore betting at MP would be worthless. This is is true under the assumption that information regarding a certain horse can be revealed only once.

$$\left(\frac{W_n^{MP}}{W_n}\right) - \left(\frac{W_n^{MP}}{W_n}\right) \frac{w_{n,j}^{MP}}{W_n^{MP}} \frac{1}{\phi_j(\tau)} = \frac{1}{N}\left(\frac{W_i}{W_n}\right) \frac{E_\tau\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(\tau)}\right),0\right\}}{E_\tau\left[P_j(T)\right]} \tag{2.11}$$

where the term $\pi_j^{MP} = \frac{w_{n,j}^{MP}}{W_n^{MP}}$ captures outsiders' new subjective probabilities as these have been shaped after observing the insider trading pattern at stage 1 and $\omega_j^{MP} = \frac{W_n^{MP}}{W_n}$ is the proportion of outsiders that bets at this second stage.

Price updating effectively continues until the last stage at which starting prices (SP) are determined as the equilibrium prices observed in the market at the end of betting. Under the assumption of zero bookmaker profit, the sum of SP would be equal to one. Then, following our model, in the presence of insiders the sum of OP should always be greater in any race than the sum of SP. In reality, the sum of OP is always greater in any race than the sum of SP even in the apparent absence of insider trading.[7] The reason is that opening prices tend to have a "cartel" level of profit built in since they are recommended to individual bookmakers by the bookmakers' association. Once betting begins, there is competition among bookmakers and thus the sum of prices will tend to decrease. This practically means that the estimates of insider trading obtained when applying our model may overestimate its true extent if the premium included in OP is largely due to this "cartel" profit rather than to the risk that bookmakers face in the presence of insiders. On the other hand, it may be that the expected profit margins built into OP are designed just to compensate the bookmakers for inside trades.

2.3.2 The option analogy

The commitment made by bookmakers to sell at fixed prices, the quoted odds, can be analyzed as a call option. Specifically, the bookmaker gives an insider a call option on horse j, i.e., the right to bet at a fixed price. Obviously, the underlying asset whose value changes stochastically is horse's j true winning probability. Apparently, only insiders are entitled to the option. The reason is that while an insider has perfect

[7] Races in which there are no plunges visible in the data (odds at no point fall for any horse during the betting) are races in which inside trades are not observed. Of course, it could be that an insider has placed a discreet bet with a single bookmaker and that this bet cannot be discerned in the average odds that rule in the market and are published. The greater the extent of this phenomenon, the more will our estimates of insider trading underestimate its true extent.

information (both public and private) and therefore knows her horse's true winning probability, outsiders form their expectations, at least partially, according to the public component of information and based on that they assign subjective probabilities. The insider will exercise her option to bet at her horse at the opening prices only if she expects a positive return, i.e. if the true probability at the time the insider places her bet, t^*, is greater than the opening price.

One could assume that insiders would be better off exercising their option at the last minute, i.e. at the closing of betting. The reason is that since a plunge reduces the prices of other horses, each group of insiders has an incentive to wait for the other groups to plunge first. However, since the insider information concerning any given horse is likely known to more than one person, the longer insiders wait, the greater the risk that the information will leak to a third party. The recipient of the leak will then plunge the horse and the group of insiders - except perhaps the one responsible for the leak - may be left with odds at which betting is no longer worthwhile. This timing dilemma is similar to the problem of the optimal exercise time faced by the holder of an American option on a dividend-paying stock. In the betting market, the dividend equivalent is the potential value leakage as a result of competition among insiders. However, for simplicity we ignore competitive interactions among insiders. We assume instead that insiders place their bet once they observe mispricing.

Assuming that the bookmaker is risk-neutral then today's option price (time zero) can be determined by discounting the expected value of the terminal option price by the riskless rate of interest. Therefore, neglecting the time-value of money, the value of the call option is:

$$C_j(0) = C_j^{OP} = E_0 \left\{ \max \left(-1 + \frac{P_j(T)}{\phi_j(0)} \right), 0 \right\} \tag{2.12}$$

Similarly:

$$C_j(\tau) = C_j^{MP} = E_\tau \left\{ \max \left(-1 + \frac{P_j(T)}{\phi_j(\tau)} \right), 0 \right\} \tag{2.13}$$

The value of the option can be derived by assuming a stochastic process for the true winning probability and performing Monte Carlo simulations (see Section 4).

2.3.3 The favorite-longshot bias

In this section we show that the optimal prices set by the bookmaker using Equation 2.9 will exhibit the favorite-longshot bias.

Expected returns will exhibit the favorite-longshot bias if and only if $\frac{\partial E(R_j)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0$, where $E(R_j) = -1 + \frac{E_0\left[P_j(T)\right]}{\phi_j}$. This is equivalent to:

$$\frac{\partial\left(\frac{E_0\left[P_j(T)\right]}{\phi_j}\right)}{\partial E_0\left[P_j(T)\right]} > 0 \tag{2.14}$$

Denoting $f_j(0) = \frac{E_0\left[P_j(T)\right]}{\phi_j}$, Equation 2.9 can be written:

$$\left(\frac{W_n^{OP}}{W_n}\right) E_0\left[P_j(T)\right] - \left(\frac{W_n^{OP}}{W_n}\right)\left(\frac{W_{n,j}^{OP}}{W_n^{OP}}\right) f_j(0) = \frac{1}{N}\left(\frac{W_i}{W_n}\right) E_0\left\{\max\left(-1+f_j(T)\right),0\right\} \tag{2.15}$$

where $f_j(T) = \left(\frac{E_T\left[P_j(T)\right]}{\phi_j}\right) = \frac{P_j(T)}{\phi_j}$.

Differentiating the above with respect to $E_0\left[P_j(T)\right]$ and setting $\frac{\partial E_0\left\{\max\left(-1+f_j(T)\right),0\right\}}{\partial E_0\left[P_j(T)\right]} = \frac{\partial E_0\left\{\max\left(-1+f_j(T)\right),0\right\}}{\partial f_j(0)}\frac{\partial f_j(0)}{\partial E_0\left[P_j(T)\right]}$, we obtain:

$$\frac{\partial f_j(0)}{\partial E_0\left[P_j(T)\right]} = \frac{\left(\frac{W_n^{OP}}{W_n}\right)\left(1 - f_j(0)\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)}\right)}{\left(\frac{W_n^{OP}}{W_n}\right)\left(\frac{w_{n,j}^{OP}}{W_n^{OP}}\right) + \frac{1}{N}\frac{W_i}{W_n}\frac{\partial E_0\left\{\max\left(-1+f_j(T)\right),0\right\}}{\partial f_j(0)}} \tag{2.16}$$

We focus on the denominator first. The first term is always positive. The second term is positive too since the partial derivative $\frac{\partial E_0\left\{\max\left(-1+f_j(T)\right),0\right\}}{\partial f_j(0)}$ is always positive. Note that a higher level of $f_j(0) = \left(\frac{E_0\left[P_j(T)\right]}{\phi_j}\right)$ is equivalent to a lower quoted price for the same level of expected true probability. Therefore, the potential profit of insiders, as captured by the term $E_0\left\{\max\left(-1+f_j(T)\right),0\right\}$, should increase since a lower quoted price makes underpricing more likely. In the terminology of options this is equivalent to saying that a lower strike price increases the value of a call option.

We turn our attention now to the nominator. For the nominator to be positive, it is necessary that the term $1 - f_j(0)\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)}$ is positive. This obviously depends on the partial derivative of the outsiders' subjective probability with respect to the expected true winning probability. Suppose that outsiders tend to overestimate the winning chances of longshots relative to those of favorites, as often argued in the literature, i.e. $\frac{\partial\left(E_0\left[P_j(T)\right]\right)/\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0 \Leftrightarrow 1 - \frac{\left(E_0\left[P_j(T)\right]\right)}{\left(w_{n,j}^{OP}/W_n^{OP}\right)}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0.$

We know that:

$$\phi_j > \frac{w_{n,j}^{OP}}{W_n^{OP}} \Rightarrow \frac{1}{\phi_j} < \frac{1}{W_{n,j}^{OP}/W_n^{OP}} \Rightarrow \frac{E_0\left[P_j(T)\right]}{\phi_j} < \frac{E_0\left[P_j(T)\right]}{W_{n,j}^{OP}/W_n^{OP}}$$

$$\Rightarrow \frac{E_0\left[P_j(T)\right]}{\phi_j}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} < \frac{\partial\left(E_0\left[P_j(T)\right]\right)}{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)}$$

$$\Rightarrow 1 - \frac{E_0\left[P_j(T)\right]}{\phi_j}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 1 - \frac{\partial\left(E_0\left[P_j(T)\right]\right)}{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}\frac{\partial\left(w_{n,j}^{OP}/W_n^{OP}\right)}{\partial\left(E_0\left[P_j(T)\right]\right)} > 0$$

Therefore, we have proved that when the bookmaker sets optimal prices following our model, expected returns will exhibit the favorite-longshot bias provided that either outsiders have no biases in their expectations and therefore their subjective probabilities reflect the publicly available information or that they tend to overestimate the winning chances of longshots relative to those of favorites.

2.4 Empirical model

2.4.1 Option-pricing specifications of the model

The challenge faced here is that the assumed specification must be a realistic description of probability dynamics. In particular, we want to model the true winning probability such that the following requirements are met: Firstly, said probability is concentrated on $[0, 1)$. A probability of a certain horse winning equal to one implies that the probabilities of all other horses are zero. In practice this is never the case. For this reason, we set as an upper boundary for the true winning probabilities the value $p_{\max} < 1$. In particular, $p_{\max}$ could be the highest single probability in our sample, which is found to be 0.7197. Secondly, the sum of probabilities is equal to one at all times. Thirdly, it may exhibit positive and/or negative jumps throughout the betting period following the release of new private information. Finally, in the long-run it reverts to a mean equal to the reciprocal of the number of runners in a race. This assumes that over a long period of time all horses have equal chances of winning. Note that the behavior of this process in the absence of mean reversion is problematic since in this case the boundaries become absorbing.

Taking the above under consideration, the following stochastic process is assumed:

$$dP_j(t) = h\left(\mu - P_j(t)\right)dt + P_j(t)\left(p_{\max} - P_j(t)\right)Jdq \tag{2.17}$$

where h is the speed of mean reversion, μ is the long-run mean (equal to $\frac{1}{N}$), J is the jump size which is assumed to be normally distributed with mean zero and standard deviation σ_j and dq describes a time-homogeneous Poisson jump process such that $dq = 1$ with probability λdt and $dq = 0$ with probability $(1 - \lambda dt)$. Parameter λ is known as intensity or arrival rate and is the expected number of "events" or "arrivals" that occur per unit time. The term $P_j(t)\left(p_{\max} - P_j(t)\right)$, which multiplies the jump component Jdq, is employed in order to ensure that the probability will remain inside the boundaries of zero and $p_{\max}$. Furthermore, given that the jump size has a mean of zero, it can be easily shown that the expected value of $P_j(t)$ at any $t > 0$ is given by:

$$E\left[P_j(t)\right] = P_j(0)e^{-ht} + \mu\left(1 - e^{-ht}\right) \tag{2.18}$$

Note that when the speed of mean reversion is very small, as assumed in this chapter, the expected value of $P_j(t)$, $E\left[P_j(t)\right]$, tends to the initial value $P_j(0)$. This is important since the theoretical model described previously in this chapter relied heavily on this assumption.

There is one final concern with respect to the specification for the true winning probability, which, as mentioned earlier, refers to the fact that the sum of probabilities must be equal to one at all times. Suppose that the probability of horse k follows the above stochastic process, while for all other horses $j, j = 1, 2..., N, j \neq k$, it holds that:

$$dP_j(t) = h\left(\mu - P_j(t)\right) + \epsilon_j(t) \tag{2.19}$$

Then, taking the sum of all probabilities, setting it equal to one and observing that

$$\sum_{j=1}^{N} h\left(\mu - P_j(t)\right) dt = 0,$$

it follows directly that:

$$\sum_{\substack{j=1 \\ j \neq k}}^{N} \epsilon_j(t) + J_k P_k(t) \left(p_{\max} - P_k(t)\right) dq = 0 \tag{2.20}$$

Therefore, although Equation 2.17 does not warrant that $\sum_{j=1}^{N} P_j(t) = 1$, we can find a condition under which this holds. Thus the above specification is indeed a realistic description of probability dynamics. We now need to estimate the parameters that appear in the stochastic process followed by the true winning probability. For the purpose of this estimation we will ignore the mean-reverting component, assuming instead that the speed of mean reversion is very close to zero. Thus, we only have to estimate the parameters of the jump process and in particular, the standard deviation σ_j of the jump size and the intensity λ of the Poisson process. The intensity parameter tells us how often the true winning probability experiences a sudden jump, while the parameter of jump volatility measures the size of these jumps. We calculate these parameters by computing the second and fourth (raw) moments. These are specified as following:

$$\mu_2 = E\left(Y^2\right) = E\left(J^2\right) E\left(dq^2\right) = \sigma_j^2 \lambda \Delta t \tag{2.21}$$

$$\mu_4 = E\left(Y^4\right) = E\left(J^4\right) E\left(dq^4\right) = 3\sigma_j^4 \lambda \Delta t \tag{2.22}$$

where $Y = \frac{\Delta P}{P(1-P)}$.

Those two moments completely identify the jump components. Moreover, they can be derived from the bookmakers' odds as following: As the dataset includes prices at three points in time (*OP*, *MP* and *SP*), prices are available roughly every 15 minutes. The fifteen-minute moments may thus be calculated for each race:

$$2M : m_2 = \frac{1}{s-1}\left(u_1 - u_2\right)^2 \tag{2.23}$$

$$4M : m_4 = \frac{1}{s-1}\left(u_1 - u_2\right)^4 \tag{2.24}$$

where $s = 2$ and $u1, u2$ can be calculated as following:

$$u1 = \frac{\phi^{MP} - \phi^{OP}}{\phi^{OP}(1 - \phi^{OP})} \tag{2.25}$$

$$u2 = \frac{\phi^{SP} - \phi^{MP}}{\phi^{MP}(1 - \phi^{MP})} \tag{2.26}$$

Obviously, the one-minute moments can be calculated from the fifteen-minute moments by dividing with fifteen. Using Equations 2.21 and 2.22 for $\Delta t = 1$ minute and the estimated values for the one-minute moments, we determine the jump components σ_j and λ for all the horses in each race:

$$\lambda = \frac{3\mu_2^2}{\mu_4} \tag{2.27}$$

$$\sigma_j = \sqrt{\frac{\mu_4}{3\mu_2}} \tag{2.28}$$

Finally, we calculate the average values of σ_j and λ for our sample, which then are used in the options calculations. Note that these are "one-minute" values. For example $\lambda = 0.1$ implies that we have a jump every 10 minutes. The results are presented below.

2.4.2 A measure of insider trading

We focus now on the task of estimating the degree of insider trading, i.e. the parameter $\frac{W_i}{W_n}$. To this end, we assume that in practice bookmakers set their prices according to the methodology described above. Thus, using the actual prices, we can infer the degree of insider trading by using Equation 2.6 to directly solve for $\frac{W_i}{W_n}$. However, the theoretical model was built under the assumptions of zero expected profit and zero transaction costs. This may yield estimates of insider trading that are biased upward. Starting from Equation 2.6:

$$1-\sum_{j=1}^{N}\frac{w_{n,j}^{OP}}{W_n^{OP}}\frac{E_0\left[P_j(t)\right]}{\phi_j(0)}=\frac{1}{N}\frac{W_i}{W_n^{OP}}\sum_{j=1}^{N}E_0\left\{\max\left(-1+\frac{P_j(T)}{\phi_j(0)}\right),0\right\} \tag{2.6}$$

By multiplying with $\frac{W_n^{OP}}{W_n}$, the part of noise trading that occurs at OP:

$$\frac{W_i}{W_n}\frac{1}{N}\sum_{j=1}^{N}C_j^{OP}=\frac{W_n^{OP}}{W_n}\left(1-\sum_{j=1}^{N}\frac{w_{n,j}^{OP}}{W_n^{OP}}\frac{E_0\left[P_j(t)\right]}{\phi_j(0)}\right) \tag{2.29}$$

or that

$$q\sum_{j=1}^{N}C_j^{OP}=N\frac{W_n^{OP}}{W_n}D^{OP} \tag{2.30}$$

Where $D^{OP}=\left(1-\sum_{j=1}^{N}\pi_j^{OP}\frac{E_0\left[P_j(t)\right]}{\phi_j(0)}\right)$ for each race. The superscript OP indicates that these values refer to the first stage at which the opening prices are set. This is the basic equation for our empirical analysis. Obviously this expression refers only to OP. However, as we said, we assume that trading takes place in two stages. At stage 1 ($t_1=0$), a proportion of the outsiders bet in the market at the OP set by the bookmaker. At any subsequent point in time, $t_1+\Delta t\leq T$ all insiders may bet should the opportunity arise. The bookmaker observes the insider trading pattern and at time, $t_2, t_1<t_2\leq T$ updates his prices. At stage 2, the rest of the outsiders bet at the new set of updated prices, denoted by MP. Again, at any subsequent point in time, $t_2+\Delta t\leq T$ all insiders may bet should the opportunity arise. The bookmaker observes the insider trading pattern and at time T sets new updated prices denoted by SP (starting prices).

Similarly, for the second stage at which MP are set, we have:

$$q\sum_{j=1}^{N}C_j^{MP}=N\frac{W_n^{MP}}{W_n}D^{MP} \tag{2.31}$$

where $D^{MP}=\left(1-\sum_{j=1}^{N}\pi_j^{MP}\frac{E_\tau\left[P_j(\tau)\right]}{\phi_j(\tau)}\right)$.

We can use Equations 2.30 and 2.31 to calculate the proportion of outsiders that bet at OP and MP:

$$\omega^{OP} = \frac{W_n^{OP}}{W_n} = \frac{D^{MP} \sum_{j=1}^{N} C_j^{OP}}{D^{MP} \sum_{j=1}^{N} C_j^{OP} + D^{OP} \sum_{j=1}^{N} C_j^{MP}} \tag{2.32}$$

Then we can use Equations 2.30 and 2.32 to calculate q. In order to do so we still have to explain how to calculate the option values at both OP and MP, C_j^{OP} and C_j^{MP}, as well as the quantities D^{OP} and D^{MP} for each race.

We begin with betting at OP. The option values C_j^{OP} can be estimated via Monte Carlo simulation. The required inputs to perform the simulations are $P_j(0)$, OP_j, T and the specifications of the stochastic process followed by the true winning probability. OP_j is the observed opening price quoted by the bookmaker. T is assumed to be equal to 30 minutes. With respect to the specifications of the stochastic process followed by the true winning probability, we need to know the speed of mean reversion, h, the long-run mean, μ, which is set equal to $\frac{1}{N}$, the standard deviation of the jump size, σ_j, given that $J \sim N\left(0, \sigma_j\right)$ and the intensity λ of the Poisson process. A way to derive those parameters has been shown in the previous section of the chapter. The speed of mean reversion is assumed to be very small since we are dealing with a betting period of no more than 30 minutes, and is therefore set at 0.001. The true winning probability, $P_j(t)$, is derived via a conditional logit regression on a dummy win, ensuring that the sum of probabilities in each race equals 1. The subjective probabilities π_j^{OP} are calculated by simply normalizing OP as suggested by Dowie (1976), although this yields estimates with a favorite-longshot bias. The true winning probability is simulated in 1000 steps using the stochastic process in Equation 2.17. When the simulated true winning probability after 1000 steps is larger than the true winning probability in time $t = 0$, the option value is this difference, otherwise the option value is zero. Each horse is subject to 1000 repetitions. The option value for the horse is the averaged value over all repetitions. A similar procedure is followed to calculate C_j^{MP}, using as inputs $P_j(\tau)$, MP_j and $T - \tau$, where τ is assumed to be equal to 15 minutes. We use the same specifications for the stochastic process as above.

We still need to calculate D^{OP} and D^{MP} for each race. The expected true winning probabilities at the end of the betting period, $E_0\left[P_j(T)\right]$, are assumed to be equal to the true winning probabilities at time 0, i.e. $E_0\left[P_j(T)\right] = P_j(0)$. This is derived from Equation 2.18 if we assume that the speed of mean reversion is very small. This way we assume that mean reversion has almost no effect on the true winning probabilities in the very short betting period of 30 minutes, while any deviations from the initial value are due to the effect of jumps that come as surprises.

Next, we use Equation 2.32 to calculate $\omega^{OP} = \frac{W_n^{OP}}{W_n}$. The extent of insider trading for each race is then:

$$q = \frac{N\omega^{OP}D^{OP}}{\sum_{j=1}^{N} C_j^{OP}}$$

The probability of insider trading is simply:

$$a = \frac{q}{1+q}$$

As we said, our model is built from the viewpoint of the bookmaker and the approach we have followed so far effectively supposes that bookmakers know the probability of insider trading in advance. Or, more reasonably, such a measure is of the bookmakers' expectations regarding insider trading. However, we have access to *ex-post* plunging information which the bookmaker cannot know until after insider activity has taken place. We will use this (*ex-post*) plunging information in order to get closer to the true probability of insider trading for a given horse that got plunged.

In order to calculate the probability of insider trading per horse, we use both an unweighted and a weighted average of q. The weight is derived as the absolute size of the plunge called PW: $\max(MP - OP, 0) + \max(SP - MP, 0)$. Using the unweighted and weighted average, the probability of insider trading for each horse in a given race in the sample is calculated. Note that when we weight absolute plunges sizes, we are weighting on those horses where insiders were observed to have bet more heavily in accordance with plunge size. Using these weights, the weighted average probability of insider trading for each of the races in the sample is calculated. The simple average of these values is the probability of insider trading in the dataset.

Table 2.1: Conditional logit regression

	Win	Win
OP	6.713***	
	(0.133)	
MP		7.155***
		(0.141)
N	45,266	45,266
Log Likelihood	-8,259.18	-8,238.81

Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

2.5 Results

We use the above model to derive a measure of the extent of insider trading. Our measure is applied to a dataset of the 1998 Australian Horse Racing season, covering 4,017 races with 45,296 runners.[8] The dataset includes for each horse prices at three moments (*OP*, *MP* and *SP*). The time period during which betting takes place is set at 30 minutes, meaning that prices are available roughly every 15 minutes. Given that there were some cases in which the sum of *OP*, *MP* or *SP* in a race is less than one, these races are dropped from the sample. This leaves us with 3,995 races out of the initial sample of 4,017 races.

The true winning probabilities at *OP* and *MP*, necessary for the measure, are derived by running a conditional logit regression, the results of which are displayed in Table 2.1.

Table 2.2 displays descriptive statistics for *OP*, *MP*, *SP*; the subjective winning probabilities at *OP* and *MP*; the true winning probabilities flowing from Table 2.1; and the sum of *OP*, *MP* and *SP* per race.

The table shows clearly that the average sum of prices decreases between *OP* and *SP*. At the opening of betting this margin is 43 percent, but by the start of the race the margin has decreased to 24 percent. The decrease in the margin indicates competition among bookmakers, forcing them to decrease prices and leading to lower profits. Since the OP are above the competitive level, this could deter insiders from trading at these

[8]The data were obtained from the CD-Rom, *Australasian Racing Encyclopedia '98*, presented by John Russell.

Table 2.2: Descriptive statistics

Variable	Mean	Min	Max	St. Dev
OP	0.1255	0.0019	0.8889	0.1017
MP	0.1107	0.0013	0.867	0.0946
SP	0.1092	0.001	0.8462	0.0972
$\pi_j(OP)$	0.0883	0.0014	0.7197	0.0731
$\pi_j(MP)$	0.0883	0.0011	0.7165	0.0768
$P_j(OP)$	0.0883	0.0026	0.9657	0.097
$P_j(MP)$	0.0883	0.002	0.9723	0.0984
$\sum_{j=1}^{N} OP$	1.4339	1.0225	2.0631	0.1117
$\sum_{j=1}^{N} MP$	1.266	1.0003	1.8508	0.1008
$\sum_{j=1}^{N} SP$	1.2487	1.0122	1.7646	0.0921

Table 2.3: Option statistics

Variable	N	Mean	Max	St. Dev
C_j^{OP}	40082	0.00454	0.20610	0.01385
C_j^{MP}	38460	0.00723	0.25498	0.02087

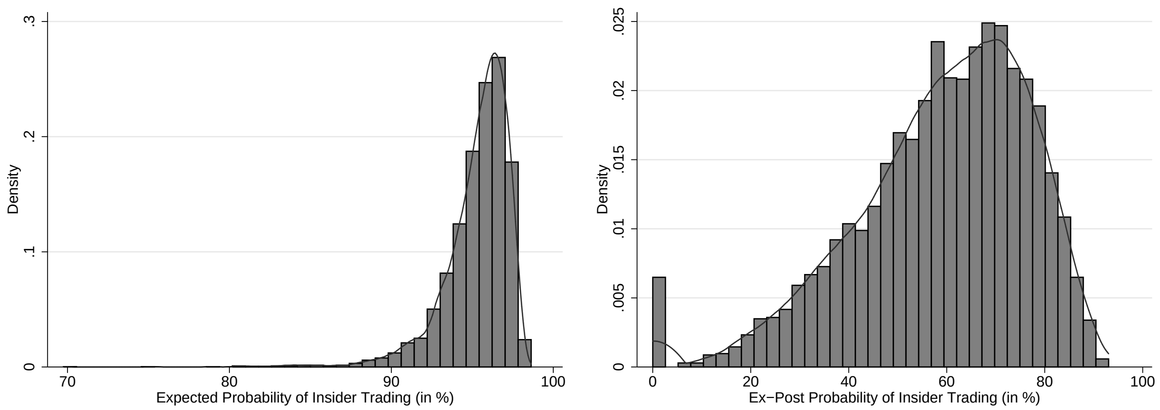
prices, leading to a lower degree of insider trading.

The option values are generated via Monte Carlo simulation as explained in the previous section. The average 1 minute values for λ and σ in the dataset are 0.37 and 0.11 respectively. On average, there seems to be a jump every 3 minutes, or 10 times per a 30 minute betting period, indicating there is quite some inflow of private information into the prices. Table 2.3 shows the values of the non-zero options generated at OP and MP.

There were 5,184 horses for which a zero option was generated at OP and 6,806 horses for which the option value was zero at MP. Moreover, we can see from Table 2.3 that the options generated at MP have a higher value, indicating more profitable trading opportunities for insiders. This should not come as a surprise as we already saw in Table 2.2 that the MP are lower, leading to a lower strike price for the insiders and a higher profit.

Figure 2.1: Density plots

(a) Expected probability of insider trading **(b)** Ex-post probability of insider trading



One last thing is required to calculate the degree of insider trading, namely ω^{OP}, the part of outsiders that bet at opening prices. Using the data from Tables 2.2 and 2.3, the average ω^{OP} in the dataset is calculated to be 0.41. On average 41% of outsider trading occurs at OP and 59% at MP, although there are races in which almost no outsider betting is found to occur at OP. The ratio of insider betting to outsider betting, which is expected by the zero profit bookmaker is found to have a mean of roughly 25, or an average probability of insider trading of around 95%. The density plot is shown in Figure 2.1a.

This seems very high but there are a few considerations to take into account. First, the insider in our model is assumed to know only the true winning probability of one horse k and does not know the winning probabilities of any other horses. When compared to the insider by Shin (1991, 1992, 1993), our insider does not know which horse will win but just has a better understanding of the true winning probabilities compared to the probabilities as quoted by the bookmaker. Second, the definition of insider trading is the amount of money being bet by insiders compared to outsiders. The fact that 95% of the total money is bet by insiders, does not mean that they place more bets, but could mean that they wager more money per bet. The bulk of the bets placed could still be made by outsiders, but the amount that insiders bet compared to outsiders is just much higher, i.e. for every Australian dollar bet by outsiders 25 is bet by insiders. Third, the assumption underlying the measure is that the bookmaker set his expected profit to zero, as would be the case under perfect competition. This is, of course, a very strict assumption to make and may not suit the reality all that well. A solution would be to assume that the bookmaker sets prices that guarantee him a certain level of profit. This level could be assumed to be equal to the profit the bookmaker would make in a market with no insiders. However, the prices that the bookmaker would set in this market are unobservable by definition. By looking at Equation 2.5, keeping everything else constant and assuming the bookmaker sets his prices to have a positive expected profit, it becomes obvious that we are estimating the degree of insider trading with an upward bias.

Finally, the measure that is generated is the bookmaker's expectation regarding insider trading. To have a zero expected profit, the bookmaker expects the probability of insider trading to be 95%. However, as mentioned in Section 4, we will use *ex-post* plunging information to get closer to the true probability of insider trading in the dataset. The weight that we use is based on the absolute size of the plunge: $PW = \max(MP - OP, 0) + \max(SP - MP, 0)$. We use PW to weigh the extent of insider trading, q, per race. By defining a plunge as an upward movement in the price, we find that there have been 13,852 plunges in the dataset, mainly occurring between MP and SP. The 13,852 horses account for 30% of the total observations in the dataset. An additional benefit is that PW weights horses that have experienced a more severe plunge higher. However, a downside is that the estimate will be too high if part of the plunging is actually due to herding. The mean of the weighted extent of insider trading q_{PW} for the dataset is 2.20 and the average probability of insider trading a_{PW} is 59%. Figure 2.1b displays the distribution of the ex-post probability of insider trading.

We can see that there are around 611 races in which no plunges occur and hence no insider trading is observed. When insider trading does take place, the average probability is around 60%, although there is plenty of dispersion around the mean.

One final remark should be made with respect to the ex-post probability of insider trading. It should be noted that the value depends on the bookmaker's expectation of the degree of insiders compared to outsiders. If we allow for a higher than zero expected profit, his expectation will be lower and we would find values of insider trading closer to the 20%-30% range that Schnytzer, Lamers & Makropoulou (2010a) find.

2.6 Conclusion

In this chapter we model a fixed odds horse betting market from a bookmaker's point of view under uncertainty. We rely upon a model by Makropoulou and Markellos (2011) and Schnytzer, Lamers & Makropoulou (2010a) which conceptualizes fixed odds betting markets as option markets. Starting from a profit function, we show that a bookmaker offers an implicit call option to insiders when setting prices. The insiders in this chapter are assumed to know only the true winning probability of their horse, not the identity of the winning horse. Moreover, insiders only bet if their expected profit is positive. In the case in which both outsiders and insiders exist in the market, the bookmaker will set prices in such a way that his expected loss from dealing with insiders equals the expected gain from dealing with outsiders. When the bookmaker set his prices in this way, the latter will exhibit a favorite-longshot bias.

By allowing for betting in multiple time periods and making an assumption on how the insider money will arrive, the zero profit condition of Schnytzer, Lamers & Makropoulou (2010a) has to hold only for the race and not for each individual horse. From this model it becomes possible to measure the expectations of the bookmaker regarding the ratio of insider money to outsider money. Using Monte Carlo simulations, we generate the implicit option values as quoted by the bookmaker and find that a zero expected profit bookmaker expects 95% of the money bet to be placed by insiders. However, these estimates are biased in the sense that we do not allow the bookmaker to make a positive profit. By keeping his expected profit equal to zero, we overestimate the expected degree of insider trading. When we use ex-post plunging information, we conclude that the probability of insider trading in our dataset lies around 59%. Or to put it differently, for every Australian dollar bet by outsiders, the average amount bet by insiders is 2.20 dollars.

Appendix 2.A Proof

Suppose first that only outsiders exist in the market and that their demand is inelastic, i.e. $\frac{\partial W_n}{\partial OP_j} = 0$. A monopolistic bookmaker will set prices that maximize her expected profit:

$$\max E(\Pi) = W_n - \sum_{j=1}^{N} E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{1}{OP_j}$$

$$\frac{\partial E(\Pi)}{\partial OP_j} = 0 \Rightarrow -E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{-1}{OP_j^2} = 0$$

Since infinite prices do not make any sense, this leads to the conclusion that outsiders' demand should be elastic, i.e. $\frac{\partial W_n}{\partial OP_j} < 0$. In this case we have the following solution:

$$\frac{\partial E(\Pi)}{\partial OP_j} = 0 \Rightarrow \frac{\partial W_n}{\partial OP_j} - E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{-1}{OP_j^2} - E\left[P_j(T)\right] \frac{w_j}{W_n} \frac{1}{OP_j} \frac{\partial W_n}{\partial OP_j} = 0$$

Suppose now that insiders also exist in the market:

$$\max E(\Pi) = W_n \left(\sum_{j=1}^{N} OP_j \right) - \sum_{j=1}^{N} E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{1}{OP_j} - W_i \max_j C_j$$

$$\frac{\partial E(\Pi)}{\partial OP_j} = 0 \Rightarrow$$

$$\frac{\partial W_n}{\partial OP_j} - E\left[P_j(T)\right] \frac{w_j}{W_n} W_n \frac{-1}{OP_j^2} - E\left[P_j(T)\right] \frac{w_j}{W_n} \frac{1}{OP_j} \frac{\partial W_n}{\partial OP_j} - \frac{\partial W_i}{\partial OP_j} \max_j C_j - W_i \frac{\partial C_i}{\partial OP_j} = 0$$

In the above equation all terms except for the first one are positive. Specifically, $\frac{\partial W_i}{\partial OP_j}$ is negative given that insider demand drops as the price increases and $\frac{\partial C_i}{\partial OP_j}$ is negative since the option price decreases as the strike price increases (or equivalently as the level of moneyness decreases).

3

The Impact of Insider Trading on Forecasting in a Bookmakers' Horse Betting Market*

3.1 Introduction

The successful forecasting of horse race outcomes requires the forecaster to have a clear understanding of the variables at his disposal. The most common, and arguably the most important, variables in a horse betting market are the odds on the horses in a race. Where bookmakers operate in such a market, it seems reasonable to suppose that the fixed odds they provide would be reasonably unbiased estimators of the horses' winning probabilities, and yet there is a considerable body of literature suggesting that this is not so (see, for example Schnytzer and Shilony, 2003; Shin, 1991, 1992, 1993). It is agreed that the extent of insider trading in the market is what makes bookmakers' odds deviate from winning probabilities, even though different authors characterize both the mechanism underlying the concomitant distortion and its extent differently.

*This chapter is based on Schnytzer, Lamers & Makropoulou (2010b), published in the *International Journal of Forecasting*.

Accordingly, the forecasting of race outcomes should take into account an estimate of both the extent of insider trading for each horse, and the way in which this extent of insider trading in a bookmakers' horse betting market may be measured. Schnytzer, Lamers & Makropoulou (2010a), SLM hereafter developed a model for measuring the extent of insider trading in horse betting markets with bookmakers.[1] They develop a theoretical framework that examines the optimal price setting by bookmakers in the racetrack betting market, and then use it to measure the extent of insider trading in the market. Bookmakers are faced with the risk that insiders will account for information they might have after the opening odds (which may be assumed to contain most public information) have been set, and will thus exploit any mis-pricing by the bookmaker by betting on horses whose prices present an expected profit for the insider. The model is an extension of that developed by Makropoulou and Markellos (2011) and applied to the European soccer betting market. The basic intuition underlying the model is that the fixed odds[2] offered by bookmakers at the track are examples of call options, and that, while bookmakers hope to offer only net of premium out-of-the-money options, when they err by underestimating a particular horse's true winning probability, they are liable to offer a net in-the-money option, which the insider (who is assumed to know her horse's true winning probability) will be glad to snap up.

Throughout this chapter, we use the working assumption that the insider knows her horse's true winning probability, and this requires some elaboration. Indeed, it is difficult to come up with a precise definition of an insider trade for which data may ever be available. Thus, in reality, an insider is one who is more familiar with her horse than others, and therefore has an informational edge over outsiders, and, *ceteris paribus*, is in a better position to evaluate the horse's winning probability. However, *ceteris* is not *paribus*! There are optimistic and pessimistic insiders, just as there are different kinds of outsiders. Some people know more about forecasting and some less, and these kinds of differences are never measurable in the kinds of data sets that are available from horse betting markets. This is why we make the assumptions we do about insiders, while, with respect to outsiders, we assume that they bet according to

[1] Theirs is not the first such method. Shin (1993) developed a similar method using a very different model.

[2] For the purpose of this chapter, by odds, we mean that odds of, say, 5 to 1 represent a net profit of $5 for every $1 bet on the winning horse.

the opening odds set by bookmakers, these being the best available estimate of public information prior to the start of betting at the track.

As described in Chapter 2, a bet by insiders involves a plunge, where several agents of the insider approach different bookmakers simultaneously and back the same horse at the best odds available. For our purposes, any fall in odds (increase in price) is taken as evidence of a plunge, and we use this information as well as the SLM measures as a predictor for the outcome of the race.

We proceed as follows: in Section 3.2, the data are described and we provide a brief discussion of our forecasting method. The results are presented in Section 3.3, where it is shown that forecasting on the basis of opening prices only-the prices are readily available around 30 min before the race-yields moderate losses. The extent of these losses is reduced when the variable measuring insider trading is added, but the method employed here would be difficult, if not impossible, to implement in practice. Finally, some conclusions are offered in Section 3.4.

3.2 Data and methodology

The data set used in this chapter contains the 45,296 horses who ran in 4,017 races during the 1997-1998 Australian horse racing season. The data include opening prices (hereafter *OP*), as set by bookmakers at the start of betting (around 30 min before each race), and middle prices (hereafter *MP*), which are prices usually, but not always, set when there is a change in the direction of the horses' odds between *OP* and the odds at the end of the betting. Finally, we have starting prices (*SP*), the ruling prices at the end of betting. The data set contains all races for which *MP* are provided. The data were obtained from the CD version of the *Australasian Racing Encyclopedia '98*.

SLM estimated several alternative measures of the extent of insider trading in this market, and we use three of the estimates for our forecasting purposes. However, in order to facilitate an understanding of these measures and the differences between them, a summary of the SLM estimation procedure is in order. Bookmakers' odds, as initially set (i.e. *OP*), may be viewed as call options which end in-the-money if the horse wins the race and out-of-the-money otherwise. As inside information enters the market, the odds change and the values of the call options change. As the betting continues, the horses' winning probabilities (as implied by the odds) become more and more accurate until all inside information has entered the market and the betting comes to an end. Assuming that the inside information enters the market randomly from the point of view of the bookmakers, the dynamics underlying the changing implied winning probabilities may be modeled as a standard Wiener process.

Using a Monte Carlo simulation, we are able to derive the option value for each horse. The true winning probability for each horse is simulated in 1,000 time steps using a standard Wiener process. When the simulated probability is larger than the strike price at the 1,000th and final step, the option value is this positive difference; otherwise, the option value is zero. For each horse, the option value is calculated as the average value out of 1000 repetitions. In order to calculate the extent of insider trading, the following three weightings are used to provide us with our estimates of insider trading for the purpose of this chapter.

The first weight used for each horse is the estimated initial winning probability, as implied by OP, P(0). The remaining two weights are based on the plunge behavior[3] in the market, and are calculated as follows. The first is the relative size of the plunge, called PW : $\max((MP - OP)/OP, 0) + \max((SP - MP)/MP, 0)$. The second weight is the absolute size of the plunge, called $PW2$: $\max(MP - OP, 0) + \max(SP - MP, 0)$. Using these weights, the weighted average degree of insider trading for each of the races in the sample is calculated. The simple average of these values is the extent of insider trading in the dataset.

Table 3.1 displays the extent of plunges in the data set, where an early plunge is defined as a positive percentage price change from OP to MP and a late plunge is defined as a positive percentage price change from MP to SP. A sustained plunge is where the horse in question is subject to both early and late plunges; the extent of the sustained plunge is then the percentage change from OP to SP. It can be seen from Table 1 that the majority of the 13,852 plunges in the dataset are late plunges, suggesting insider trading at MP. However, the average extent of early plunges exceeds that of late plunges.

[3]A horse is said to have been plunged when its odds suddenly decrease meaningfully owing to large bets having been placed on the same horse with different bookmakers simultaneously. Schnytzer and Shilony (1995) show that plunges contain inside information.

Table 3.1: The extent of plunges in the dataset

Plunges	Number	Average extent (%)
Early plunge	1,281	21.25
Late plunge	9,783	15.72
Sustained plunge	2,788	26.33
All	13,852	18.37

An early plunge is defined as a positive percentage change from OP to MP. A late plunge is defined as a positive percentage change from MP to SP. When there are both early and late plunges, this is known as a sustained plunge.

Table 3.2: Measures of the degree of insider trading for each specification

Weight	Degree of insider trading (%)
$P(0) - OP$	32.68
PW	26.38
$PW2$	26.48

$P(0) - OP$ is the true winning probability at time 0. PW is $[\max((MP - OP)/OP, 0) + \max((SP - MP)/MP, 0)]$. $PW2$ is $[\max(MP - OP, 0) + \max(SP - MP, 0)]$.

The simple average of these values is a variable that measures the extent of insider trading as estimated by SLM, and is shown in Table 3.2. Armed with opening prices and various measures related to the extent of insider trading for each horse, we proceed to forecast the winners of each race in the data set. We use the generally preferred method of forecasting in the betting literature, namely the conditional logit model (hereafter CL) of McFadden (1974). We estimate several CL models. The first estimates the probability of horse i winning race j based solely on the information contained in OP, as follows:

$$p_{ij}^{o} = \exp\left(a_1 OP_{ij} / \sum_{i=1}^{n} \exp(a_1 OP_{ij})\right), \tag{3.1}$$

where $i = 1, 2, \ldots, n_j$, OP_{ij} is the OP of horse i in race j, n_j is the number of runners in race j, and $a1$ indicates the contribution which OP makes to the horse's chance of winning race j. We then run four more regressions, adding different predictors to OP in turn. These variables are as follows. First, the option value for each horse, as estimated by SLM. This variable is zero for most horses in the sample and positive for one or two in each race. Optionvalue$_{ij}$ is positive if horse i's winning probability in race j is estimated, via a Monte Carlo simulation, to be greater than the winning probability implied by OP, and is measured as the difference between the two. In this case, the model to be estimated may be written:

$$p_{ij}^{o} = \exp\left(a_2 OP_{ij} + b_1 \text{Optionvalue}_{ij}\right) / \sum_{i=1}^{n}\left(a_2 OP_{ij} + b_1 \text{Optionvalue}_{ij}\right), \tag{3.2}$$

where $i = 1, 2, \ldots, n_j$. The coefficients a_1, a_2 and b_1 in regressions (3.1) and (3.2) are measured by maximizing the joint probability of observing the winners of all of the races in the sample. Next, we add EarlyPlunge$_{ij}$, which is equal to the difference between MP and OP when this difference is positive, and zero otherwise.[4] Our fourth predictor is the extent of insider trading on horse i in race j, as measured by SLM (Insidertrading$_{ij}$). Finally, we add TotalPlunge$_{ij}$, which measures the total extent of early and late plunges on horse i in race j. We expect, a priori, that all variables should, by themselves, add to a horse's winning probability, and thus should receive positive coefficients.

3.3 Results

Table 3.3 shows the results of our five regressions. It is clear from these results that OP is by far the most important predictor of winning probabilities, in terms of both coefficient size and statistical significance. Given the bookmakers' stake in the outcome of the betting, it is clear that OP will reflect as much useful information as possible, unless bookmakers deliberately distort prices as a defense mechanism against insiders.[5] With the exception of Optionvalue, all of the variables have positive coefficients which are statistically significant in at least one of the regressions. When Optionvalue is used as the sole predictor of winning probabilities, it has a positive and highly significant coefficient,[6] leading us to conclude that the unexpected results here are the result of multicollinearity.

[4]The winning probabilities when this and subsequent variables are added may be estimated by models that follow trivially from Eqs. (3.1) and (3.2), and thus are not noted explicitly.
[5]See SLM, Schnytzer and Shilony (2003) and Shin (1991) and Shin (1992) for more discussion on this point.
[6]Full results are available upon request.

Table 3.3: Predicting horses' winning probabilities

Specification	1	2	3	4	5
Dependent variable	Win	Win	Win	Win	Win
OP	6.7117 (50.58)*	6.6583 (47.09)*	6.5204 (45.73)*	6.4404 (44.91)*	6.3347 (43.88)*
Optionvalue		2.1108 (1.09)	1.5214 (0.78)	-2.7790 (-1.32)	-13.7064 (-5.63)*
EarlyPlunge			10.824 (6.79)*	10.1933 (6.39)*	
Insidertrading				0.2741 (5.80)*	0.2079 (4.32)*
TotalPlunge					9.3020 (9.62)*
N	45,296	45,296	45,296	45,296	45,296
Pseudo-R^2	0.1389	0.1390	0.1412	0.143	0.1456

* Indicates significance at the 1% significance level. Optionvalue are the option values generated by SLM. EarlyPlunge is the extent of early plunges as measured by $\max(MP - OP, 0)$. Insidertrading is the incidence of insider trading on a specific horse, as generated by SLM. TotalPlunge is the occurrence of early and late plunges together, as measured by $\max(MP - OP, 0) + \max(SP - MP, 0)$.

Table 3.4 shows the results of betting $1 on each predicted favorite in every race in our sample, on the basis of the five regressions shown in Table 3.3. The results obtained are in line with what the regression results in Table 3.3 led us to expect. Thus, insider trading seems to influence profits (or in this case, losses) in an upward direction, although the tone is clearly set by OP, and betting on the basis of it alone leads to a loss of 10.2%. The best performance is achieved by adding option values, early plunges and the extent of insider trading to OP, but this only adds a little over 3% to the loss reduction. However, since SLM rely exclusively on price data in their simulations, these results show that this market is, in practice, weak-form efficient. Furthermore, even if the results are calculated as if betting takes place at the best odds available during the betting (as we would expect insiders to bet), rather than at SP, returns are better but remain negative throughout.[7] Finally, it may seem strange that the losses incurred in simulation 5, when all plunges are taken into account in addition to the other variables, should exceed those in simulation 4, when only early plunges are added to the model. The reason for this would appear to be the herding on late plunges in this market.[8]

[7] Full results are available upon request.
[8] See Schnytzer and Snir (2008).

Table 3.4: Betting simulations

	Betting based on the favorite predicted by specification				
	1	2	3	4	5
Number of races and bets	4017	4017	4017	4017	4017
Profit ($)	-409.83	-379.40	-360.20	-285.00	-301.50
Rate of return (%)	-10.20	-9.44	-8.97	-7.09	-7.51

Betting takes place in all races, since each race has a favorite, as measured by the highest win probability predicted after each regression specification. Betting takes place at SP, the last quoted price before the race starts.

3.4 Conclusion

In this chapter we have shown that variables which measure insider trading, as measured by SLM, have only a moderate impact on the forecasting results. Adding various different measures relating to insider trading by horse to a conditional logit model which uses only opening prices to predict winning probabilities, reduces the losses but does not generate positive profits. Therefore, the relevance of insider trading in this market, *in principle*, cannot be refuted. However, it should be pointed out that even the small gains in forecasting demonstrated here may be difficult to implement in practice.

It is unlikely that the simulations used by SLM could be carried out in the short time available before each race. Thus, a knowledge of price changes is critical, and if the latest prices used in the simulation were to be those ruling in the market five minutes or so before the start of the race, that would leave less than five minutes for the estimations. Since the simulations carried out by SLM required several days to run, a system based on our estimates could be applied only on a computer which is far more powerful than is generally available today outside the Pentagon! Furthermore, given the merely moderate gains generated by the addition of these variables to the basic model, it may be wondered whether it would be worthwhile to struggle for a solution to the computing problem.

So why has the extent of inside information not contributed more dramatically to the forecasts? To the extent that the SLM model provides a reasonable measure of the extent of insider trading, it must be concluded that the reliance on price data alone in forecasting horse races in a bookmakers' market is doomed to failure. On the other hand, perhaps the basic weakness of regression models in forecasting is that they provide predictions on the basis of “on average” results, whereas insiders bet on particular horses in particular races when as many relevant factors as possible which are unknown to outsiders have been taken into account.

4

FDI, Terrorism and U.S. Investors before and after 9/11*

4.1 Introduction

In today's globalized world, the largest and most productive companies have expanded their activities around the globe (see e.g. Melitz, 2003; Helpman, 2006). While this increased internationalization can enhance productivity and create shareholder value (Harris and Ravenscraft, 1991; Aitken and Harrison, 1999; Javorcik, 2004), it also exposes firms to new sources of geopolitical risk that threaten to destroy part of this value. The recent increase in terrorist attacks in Africa, the Middle East and parts of Asia raises the question how multinationals with investments in these regions are affected. Since investors seem remarkably apt at identifying and valuing different forms of FDI (Doukas and Travlos, 1988; Chen et al., 2000), the stock market is a potential channel through which this risk can affect the multinational.

Recently, Dube et al. (2011) and Berger and Bouwman (2013) have explored this channel using unexpected and (initially) covert CIA interventions abroad. They found

*This chapter is based on joint work with Jaap W.B. Bos (Maastricht University) and Michael Frömmel (Ghent University).

that the interventions resulted in increased exports and positive abnormal returns for U.S. multinational firms. While these interventions were generally regarded as positive for the firms involved, in this chapter we investigate whether multinationals can also see downsides as a result of geopolitical risk. We focus specifically on terrorist attacks, as the costs associated with attacks like 9/11 can range from 55 billion dollars in direct property losses, to a total estimated loss between 60 and 125 billion dollars in GDP (see e.g. Thompson Jr, 2002; Blomberg and Hess, 2009). We analyze how acts of terrorism abroad affect share prices of U.S. multinational firms that have invested in the area. According to the stock market channel, companies with investments in a region that experiences a terrorist attack see a more severe stock market reaction compared to companies that invest less in the region, or those that do not invest there at all.

Previous literature has documented that acts of terror can be transmitted through stock markets. For example, Abadie and Gardeazabal (2003) find that stocks of Spanish firms with large parts of their business in the Basque country outperformed their counterparts during a truce period, only to underperform when the truce ended. Another example is Drakos (2010a), who has found that countries with a strong trade relation with Spain and the United Kingdom experienced more pronounced abnormal stock market losses during the attacks in Madrid and London. In this chapter, we look at this relationship on a larger scale, taking into account many more instances of terrorism and their connection with FDI in Western Europe, the rest of Europe, Latin America, Africa, the Middle East and Asia.

We show that for the past several years, equity prices indeed drop after a large terrorist attack, and that the drop is related to the capital stock built by U.S. firms in the region where the attack takes place. Exploiting the fact that the 9/11 attacks take place during our sample period, we find that the relationship between large terrorist attacks, the stock of FDI built up by U.S. firms in the region and their share prices is only significant, both statistically and economically, *after* the tragic events of 9/11.

In the remainder of the chapter we look at possible explanations for this change in investor behavior, discussing both rational and behavioral factors. Based on these factors, we formulate and test several hypotheses. A first possible explanation is that terrorism in the post-9/11 world has become more frequent and violent, however we find that the total number of terrorist attacks and their severity did not change materially after the attacks. Since risk perception has been shown to be mainly attributable to acts of terrorism (Drakos and Müller, 2014), this potentially also rules out changes to the anticipation of new attacks occurring and affecting prices. The second possibility is increased media coverage, which has been shown to play an important role in the way financial markets react (Melnick and Eldor, 2010). We find that while media coverage of attacks on foreign soil initially increased after 9/11, it has since leveled out, seemingly ruling out exaggerated risk perceptions of investors as a result of media coverage (Sunstein, 2003). Third, we briefly discuss the impact of terrorism insurance in the wake of the attacks, as stock prices should theoretically not react to attacks when firms are insured against them. A short review of the literature indicates that the costs of insurance against terrorism have fallen after 9/11, and that the share of companies using terrorism insurance as guaranteed by the Terrorism Risk Insurance Act (2002) has continued to rise. Based on this information, we would expect to see a diminished relationship as time goes on. Nonetheless, the relationship between terrorist attacks, FDI and share prices is only significant after 9/11. Taken together, the evidence suggests disaster myopia of U.S. investors prior to 9/11, consistent with the availability heuristic (Tversky and Kahneman, 1973) and the concept of local thinking (Gennaioli and Shleifer, 2010), and that the attacks served as a wake-up call, creating awareness among investors about the potential impact of terrorism. These findings are in line with Malmendier and Nagel (2011), who show that extreme events are easily remembered and can have a longlasting effect on investors.

The chapter is organized as follows. Section 4.2 provides a brief literature review, while Sections 4.3 and 4.4 describe our data and methodology respectively. In Section 4.5 we present the results of the analysis before concluding briefly in Section 4.6.

4.2 Stock prices, terrorist attacks and FDI stocks

The tragic events on September 11, 2001 were the worst attacks on U.S. soil in 70 years and hit the United States in its financial and political center. The loss of lives and material damage sustained during the attacks displayed the economic consequences of terror, and the ensuing shock waves were felt around the world.

Stock market data, due to their forward looking nature and high frequency, have often been used to assess the economic impact of terrorism. Following 9/11, financial markets in America and around the world showed heavy losses as they incorporated the news of these attacks. Previous research has shown that stock markets in countries suffering from terrorism exhibit negative abnormal returns as a result of large isolated attacks like the 9/11, Madrid or London attacks (see e.g. Drakos, 2004; Carter and Simkins, 2004; Maillet and Michel, 2005). However, liquid and well diversified markets can absorb these shocks well, such that the impact of the effect is instantaneous or in some cases hardly noticeable (Johnston and Nedelescu, 2006; Mende, 2006). When taking into account a larger set of attacks, negative abnormal returns are also found (Drakos, 2010b), accompanied by an increase in stock market variance (Peren Arin et al., 2008). In an increasingly interconnected world, what remains unclear is how terrorism in different geopolitical regions can spill over to U.S. markets.

Terrorist attacks on foreign soil can affect U.S. stock prices in a number of ways relating to the discounted expected future cash flows. First, if investors experience higher risk aversion as a result of a terrorist attack, this increases the rate at which they discount future cash flows. In this case we expect to see a uniform price shock occurring across assets/sectors that share a common discount factor. However, there is ample evidence that sectors in the same country (or similar sectors across countries) exhibit different reactions to the same attack (see amongst others Chen and Siems, 2004; Straetmans et al., 2008; Berrebi and Klor, 2010; Chesney et al., 2011), making it unlikely that discount rate changes explain the price shocks we observe after terrorist attacks.

Second, an attack can influence expected future cash flows, either via an increase in expected costs or a decrease in expected revenues. Increases in expected costs can, for instance, come from increases in insurance premia or damage to the physical capital present, while decreased consumption and grown in the country where the attack took place can lead to a decrease in expected revenues (see e.g. Blomberg et al., 2004; Eckstein and Tsiddon, 2004). In both cases, the loss in cash flows depends mainly on the location of an attack, the amount of capital goods present in the location and the level of damage to these productive assets.[1] The damage, in turn, is determined by the mag-

[1] These assets can be both physical capital and human capital, although share prices react more to attacks on the latter (Karolyi and Martell, 2010).

nitude of an attack and specific factors such as population density, building codes or the quality of infrastructure. Moreover, the expected loss as a result of a terrorist attack also depends on the assessment of investors regarding the probability of attacks and their corresponding losses. Important for this assessment is how investors generate expectations. For instance, the availability heuristic can make it difficult for investors to properly incorporate all possible outcomes (Tversky and Kahneman, 1973).

Third, multinationals can also move their operations to another country as a consequence of terrorist activities. For instance, Abadie and Gardeazabal (2008) find that, in an open economy, country-specific terrorism risk can lead to movements in capital (FDI) to other countries. Enders et al. (2006), however, establish empirically that this effect is economically small and we therefore focus on the first two ways.

4.3 Data

To analyze how terrorism on foreign soil affects U.S. stock prices, we combine data from three different sources. We first describe the database on terrorist attacks, followed by data on FDI stocks and stock market data of U.S. multinational companies.

4.3.1 Terrorist attacks

We collect data on terrorist attacks from the Global Terrorism Database (GTD), developed and maintained by the University of Maryland. The GTD contains information on over 98,000 international attacks between 1970 and 2010.[2] For an event to be incorporated in the GTD, it has to be intentional, violent and carried out by non-state actors.[3] We further limit our sample by only including successful attacks that were considered to be terrorism beyond any doubt by the GTD.[4] Information on these ad-

[2]For more information on the GTD see LaFree and Dugan (2007).

[3]According to the GTD, terrorist attacks have to be aimed at 'attaining a political, economic, religious, or social goal,' there must be 'evidence of an intention to coerce, intimidate, or convey some other message to a larger audience (or audiences) than the immediate victims,' and the event must be 'outside the context of legitimate warfare activities' (Study of Terrorism and Responses to Terrorism (START), 2011). Montalvo (2011), for instance, documents how attacks indeed can attain political and religious goals by studying the outcome of the Spanish elections following the Madrid bombings, while Gassebner et al. (2011) find that terrorism shortens cabinet duration.

[4]Users of the GTD can further govern the parameters of their search results by employing an additional filter. The existence of a 'Doubt Terrorism Proper?' field records reservation, in the eyes of GTD analysts, that the incident in question is truly terrorism. Such uncertainty, however, was not deemed to

ditional filters is only available from 1998 onwards, leaving us with 16,287 attacks that take place in 144 countries on 4,009 days between 1998 and 2010. Although the number of attacks seems high, it includes both incidents of national and transnational terrorism and is in line with other papers (e.g., Piazza, 2008). Attacks that occur during weekends and holidays are placed on the next available day that the U.S. stock market can react to the attack. Out of a total 3,270 trading days in our sample, there are 2,868 during which information on terrorism can enter the market. We discuss the consequences of this high frequency of events for our analysis in Section 4.4.

In order to differentiate between the magnitude and geographical location of an act of terror, we construct a daily terrorism intensity index (Eckstein and Tsiddon, 2004; Peren Arin et al., 2008), such that

$$TER_{i,t} = \ln(1 + \#\,\text{attacks}_{i,t} + \#\,\text{fatalities}_{i,t} + \#\,\text{injured}_{i,t}), \tag{4.1}$$

where i represents the region in which the attack took place, t is the day on which the attack took place, and the number of attacks, fatalities and injuries are reported by the GTD.[5] The regions closely follow the grouping of outward FDI at our disposal and are Western Europe (comprising the European Union, Norway and Switzerland), the Rest of Europe, Latin America, Africa, the Middle East and Asia and the Pacific.

The terrorism index $TER_{i,t}$ has several attractive features. A trading day on which no terrorist attack occurs has a TER score of 0. Once attacks occur, the TER score is additive in both the number of fatalities/injuries and the number of attacks. For example, 1 attack with 10 injuries will not have the same TER score as 10 separate attacks with 1 injured person each. Whereas the former has a TER score of 2.6, the latter has a score of 3.

One downside of using the number of fatalities and injuries is that these numbers are often not known on the day of the attack itself, but can take weeks or even months to be confirmed. Even though estimated and official death tolls can differ, we assume

be sufficient to disqualify the incident from inclusion into the GTD. Furthermore, such a determination of doubt is subsequently coded by GTD analysts as conforming to one of four possible alternative designations: 1) Insurgency/Guerilla Action; 2) Internecine Conflict Action; 3) Mass Murder; or 4) Purely Criminal Act. Note that the 'Doubt Terrorism Proper' determination was only made for incidents that occurred after 1997.

[5] For the purpose of this chapter, including the intensity of media coverage in the U.S. is also a possibility, although unfortunately the GTD reports at most their top 3 sources for the information on the attack.

Table 4.1: Summary statistics of terrorism

	A. TER-scores				B. Attacks	
Region	Mean	Std. Dev.	Max.	Trading days with attack	Average deaths	Average wounded
Western Europe	0.23	0.58	7.60	640	0.43	3.90
Rest of Europe	0.42	0.93	6.98	762	2.43	5.80
Latin America	0.26	0.73	5.73	500	2.44	3.36
Africa	0.74	1.22	8.37	1,101	5.34	7.11
Middle East	1.31	1.72	6.90	1,522	4.66	11.05
Asia & the Pacific	1.74	1.63	7.57	2,086	2.77	6.16

The table displays summary statistics of attacks in different global regions selected from the Global Terrorism Database over the period 1998–2010, spanning a total of 3,270 trading days. Panel A shows the statistics of the TER scores calculated from these attacks, where 'Trading days with attacks' counts the number of days in the sample where a terrorist attack took place. Panel B shows the average number of fatalities and wounded per attack using the raw data from the Global Terrorism Database. For instance, the Madrid bombings are included as four separate attacks having injured 450 people each.

that the estimated death tolls are at least of a similar magnitude as the official death toll. By using the log transformation, we limit overestimation of the intensity score.[6]

Panel A in Table 4.1 shows summary statistics of the terrorism intensity in each of the six regions. Asia and the Pacific have experienced the highest intensity of terrorism, with an act of terrorism occurring on 2,086 out of 3,270 trading days in the sample period. This is followed by the Middle East and Africa, which experienced acts of terrorism on 1,522 and 1,101 trading days, respectively. On the contrary, these numbers are much lower for Western Europe and Latin America, where attacks only occur on 640 and 500 trading days respectively. Panel B summarizes the average number of fatalities and wounded per attack, showing that terrorism was most lethal in Africa and the Middle East. By comparison, attacks in Western Europe were the least lethal.

4.3.2 FDI stocks and stock market data

For the purpose of our analysis, we require accurate information on the FDI stock built by U.S. firms in different parts of the world. For reasons of confidentiality, such information is typically not available at the firm level. What is available from the U.S.

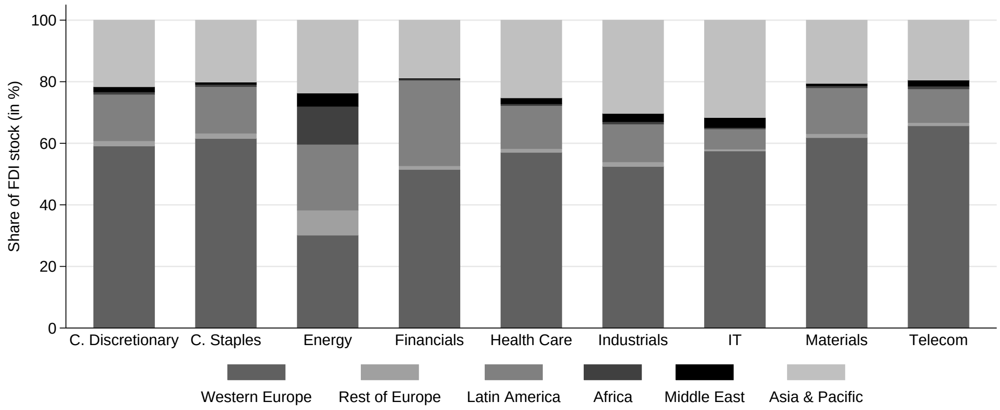
[6]For example, in the weeks after the 9/11 attacks the death toll had been estimated to be 6,000 over 4 attacks. This would have lead to a TER score of $\ln(1 + 4 + 6,000) = 8.7$. The official death toll recorded in the GTD is 2,996 (including 19 terrorists), which leads to a TER score of $\ln(1 + 4 + 2,996) = 8.0$.

Bureau of Economic Analysis (BEA), is the yearly stock of outward U.S. FDI per region and per industry, published in the 'Survey of Current Business'. The BEA covers most industries and countries, although not all combinations are reported if the amounts are negligible or if they would threaten to disclose data of individual companies. We use data for the period 1998 to 2010.

Industries are classified by the BEA using either the Standard Industry Classification (SIC) or the North American Industry Classification System (NAICS). In order to link these FDI data to stock market data, we obtain prices of the S&P500 and its subsector indices from Datastream. Since the latter are classified using the Global Industry Classification Standard (GICS), we need a mapping from SIC/NAICS to GICS, which to the best of our knowledge does not yet exist. Therefore, we obtain a list of all current and historic S&P500 companies with their SIC/NAICS and GICS codes from Compustat in order to make a conversion table. More details on the data and the conversion are provided in Appendix 4.B. In our analysis, sectors are defined according to the two-digit GICS codes and consist of Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Telecommunication Services and Utilities. For the Utilities sector, data on outward FDI were not available after 2002 and it is therefore excluded from the analysis.

Figure 4.1 shows per sector the average share of U.S. outward FDI that each region received from firms in that sector during the sample period 1998-2010. We see that Western Europe, Latin America and Asia and the Pacific all received large shares of U.S. investments. The main beneficiary of U.S. outward FDI was Western Europe, whose countries received between 50 and 70 percent of all U.S. investments made abroad. The share of Western Europe was lowest in the Energy sector, where investments were more equally distributed between the Middle East, Latin America and Asia and the Pacific. Asian and Pacific countries received their highest investment share from the IT and Industrial sectors. Moreover, outside the Energy sector, the shares of Africa and the Rest of Europe were generally quite low and averaged 0.6 percent and 1.5 percent, respectively.

Figure 4.1: U.S. Foreign Direct Investment - Sector distribution geographical areas



The figure shows the geographical distribution in outward FDI of nine S&P500 sectors. To calculate shares, data from the U.S. Bureau of Economic Analysis are transformed from NAICS/SIC to GICS sectors (see Appendix 4.B) and averaged over the sample period 1998–2010.

4.4 Methodology

Our analysis consists of three steps. In the first step, we propose a solution for the fact that terrorist attacks occur so often that we cannot make use of a standard event study methodology, since it is very difficult to construct a proper event window. In the second step, we relate share prices to terrorist attacks, allowing for the fact that investors' response may be highly non-linear, as they may only respond to attacks that are severe enough to possibly cause (future) cash flow drops. For that reason, we limit ourselves to investors' 'pure' response to terrorism in this step, and do not yet incorporate the role of FDI stocks. After all, an average attack may not have much of an effect, regardless of the FDI stock present. Finally, in the third step, we relate investor responses to large attacks with FDI stocks in different regions, taking into account the fact that this relationship may vary for reasons other than the share of total FDI each region receives.

4.4.1 First step: jumps in the price process

To identify the effect of an event like a terrorist attack, we need to separate the normal behavior of returns from abnormal behavior. The abundance of terrorist events around the globe makes it nearly impossible to find enough estimation windows with (presumably) normal returns. As this is likely to lead to biased estimates of abnormal returns when using standard event study techniques (see e.g., Craig MacKinlay, 1997), we resort to a different method.

Since terrorist attacks are typically construed to be unexpected events, we expect them to cause jumps in the price process.[7] For this reason, we rely on a jump-diffusion model to distinguish between attacks that have an impact (i.e., where a jump occurred) compared to those that do not (i.e., no jump observed). Jump-diffusion models hail back to Merton (1976), but while most of them combine a standard Brownian motion with a jump process (see e.g. Naik, 1993), we rely on a simplified GARCH-jump model proposed by Maheu and McCurdy (2004). The main advantage of this approach is that we can capture volatility clusters in a GARCH framework, something that a constant variance Brownian motion cannot do. As a result, large price changes that occur due to volatility clustering are not erroneously classified as a jump. By using this methodology, we assume implicitly that terrorism abroad will directly impact the share price in the U.S., disregarding any possible contagion effects from foreign markets to home markets (see e.g. Dungey et al., 2005; Bekaert et al., 2011). However, since we deal with many terrorist events, which tend to be absorbed quickly in liquid markets like the U.S. (Johnston and Nedelescu, 2006; Mende, 2006), we choose to look only at event days instead of possible contagion following attacks.

In order to measure jumps, we first define the standard return process. Next, we define jumps and apply a filtering procedure to separate standard price movements from jumps using a maximum likelihood estimator. Finally, we extract the probability of a jump as well as its (expected) impact on the price.

We start by describing the return process, which is defined as:

$$r_t = \mu + \phi r_{t-1} + \epsilon_t \tag{4.2a}$$

$$\epsilon_t = \epsilon_{1,t} + \epsilon_{2,t}, \tag{4.2b}$$

where r_t is the stock return, μ is a constant mean and ϕ an autoregressive component.[8]

[7] This at least holds for most of the market participants. Insider trading as reported by Poteshman (2006) should not play an important role.

[8] Adding more variables to the mean equation is possible, however our explanatory variables are indices and thus already quite broad. For example, it would not make sense to estimate a market model in this case. We have performed robustness tests where the Fama-French HML and SMB (Fama and French, 1992, 1993) factors have been included in the mean equation, but this does not change the results of the analysis. Results are available upon request.

The composite error term, ϵ_t, consists of $\epsilon_{1,t}$ and $\epsilon_{2,t}$, innovations by the GARCH and the jump process respectively. What allows us to separate $\epsilon_{1,t}$ from $\epsilon_{2,t}$ is the fact each has a different distribution, as a result of which we can decompose ϵ_t.

We begin with $\epsilon_{1,t}$, which follows a standard GARCH(1,1) process:

$$\epsilon_{1,t} \sim N(0, \sigma_t^2) \tag{4.3a}$$

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2. \tag{4.3b}$$

In contrast, $\epsilon_{2,t}$ represents the impact of jumps in the returns, which is assumed to arrive via a Poisson process with time-varying intensity:

$$\epsilon_{2,t} = J_t - E[J_t|\Phi_{t-1}], \tag{4.4}$$

where J_t is the actual jump contribution and $E[J_t|\Phi_{t-1}]$ is its expectation conditional on the previous days' returns, $\Phi_{t-1} = \{r_1, \ldots, r_{t-1}\}$. The jump contribution J_t is the sum of the stochastic number of jumps, n_t, where the size of each jump $Y_{t,k}$ is assumed to be independently drawn from a normal distribution with mean θ and variance δ^2:

$$J_t = \sum_{k=1}^{n_t} Y_{t,k}, \quad Y_{t,k} \sim N(\theta, \delta^2). \tag{4.5}$$

Finally, the probability that $n_t = j$ jumps take place on day t given the history of returns Φ_{t-1} is:

$$P(n_t = j|\Phi_{t-1}) = \frac{\exp(-\lambda_t)\lambda_t^j}{j!}, \tag{4.6}$$

where λ_t is the jump intensity. Maheu and McCurdy (2004) suggest to let the jump intensity follow an AR(1) process, where the expected number of jumps today depend on yesterday's expected number of jumps and the jump intensity residual ξ_{t-1} (the deviation of the number of jumps from its expectation):

$$\lambda_t = E[n_t|\Phi_{t-1}] = \lambda_0 + \rho E[n_{t-1}|\Phi_{t-2}] + \gamma\xi_{t-1} \tag{4.7}$$

We use the filter procedure proposed by Maheu and McCurdy (2004) and obtain the probability that at least one jump occurred based on the ex post estimation of the

Table 4.2: Summary statistics of jumps

	A. Jump probability (in %)				B. Jump contribution (in %)			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
S&P500	36.9	36.7	0.0	100	-0.7	1.2	-10.0	0.0
C. Discretionary	33.8	34.3	0.0	100	-0.7	1.2	-11.2	0.0
C. Staples	28.7	28.2	1.9	100	-0.2	0.3	-3.3	0.0
Energy	19.2	27.3	0.0	100	-0.5	1.1	-14.3	0.0
Financials	30.2	34.3	0.0	100	-0.7	1.3	-10.8	0.0
Health Care	35.1	29.9	2.8	100	-0.3	0.4	-3.6	0.0
Industrials	28.9	34.3	0.0	100	-0.7	1.3	-11.6	0.0
IT	28.5	34.0	0.0	100	-0.9	1.6	-12.5	0.0
Materials	28.8	28.0	2.3	100	-0.3	0.5	-5.1	0.0
Telecom	27.4	27.8	1.2	100	-0.2	0.3	-2.5	0.0
Utilities	13.5	24.1	0.0	100	-0.3	0.7	-11.6	0.0

The table displays summary statistics of the output from the GARCH-Jump model for the S&P500 and its sectors. Panel A shows the probability that at least one jump occurred on a trading day, $P(n_t \geq 1|\Phi_t)$. Panel B displays the jump contributions, calculated as the expected number of jumps on day t, $E[n_t|\Phi_t]$, multiplied with the average jump size θ.

number of jumps:

$$P(n_t \geq 1|\Phi_t) = 1 - P(n_t = 0|\Phi_t) \tag{4.8}$$

and the ex-post assessment of the number of jumps that occurred on each trading day:

$$E[n_t|\Phi_t] = \sum_{j=0}^{\infty} jP(n_t = j|\Phi_t) \tag{4.9}$$

Since the jump sizes are i.i.d and thus unconditional, multiplying the expected number of jumps with the average jump size yields the ex-post expected jump contribution $E[J_t|\Phi_t] = \theta E[n_t|\Phi_t]$. Together, these two elements tell us both how likely it is a jump occurred, as well as its size. Table 4.2 shows summary statistics for these two variables, based on the estimations in Table 4.A.1 in Appendix 4.A.

The sectors with the highest average jump probabilities are Health Care (35.1 percent), Consumer Discretionary (33.8 percent) and Financials (30.2 percent). In contrast, sectors with the lowest average jump probabilities are Utilities (13.5 percent) and Energy (19.2 percent). Panel B summarizes the jump contributions, and shows that while Health Care has the highest average jump probability, its largest jump contribution was only -3.6 percent.[9] The largest contribution of jumps to the Financials sector was -10 percent on August 31st, 1998, the 7th largest one-day loss on the S&P500. Sectors that have experienced both high jump probabilities and high jump contributions are Consumer Discretionary, IT and Financials.

Now that we have a measure for the likelihood and magnitude of jumps, the next question is to what extent terrorist attacks are responsible for these jumps. To investigate the relationship between terrorism, jumps and FDI, we opt for a two-stage analysis. In the first stage, we estimate per sector the relationship between the TER index and the likelihood and size of jumps. Using this estimation, we obtain the average reaction to a large attack in each of the regions. In a second stage, we regress the predicted jump probabilities and sizes on the share of outward U.S. FDI that regions received.

While we sacrifice some efficiency in our estimations by opting for this two-stage analysis, there are three important benefits to our approach. First, we consider the possibility that only the largest attacks evoke a reaction on financial markets. A single-stage analysis, with an interaction between the TER measure and FDI, would at best capture the effects of an average attack, conditional on the average FDI stock, and may thus fail to capture the conditions under which we expect to see a reaction. Moreover, since the regions experience different 'average' attacks, a two-stage analysis further allows us to obtain reactions to attacks of the same magnitude.

Second, in a single stage estimation we may fail to properly estimate the impact of terrorism conditional on FDI stock if the latter is endogenous. Indeed, Abadie and Gardeazabal (2008) find evidence that FDI can flow out of a country as a result of terrorism although empirically the effects are relatively small (Enders et al., 2006). By identifying the relationship between terrorism, FDI stocks and share prices in two steps, we are able to avoid this problem of endogeneity.

The final reason is related to the measure of outward FDI. The data obtained from the U.S. BEA are recorded in dollar amounts on a historical cost basis, and thus rise on average every year. When FDI in regions grows at more or less the same rate, only the dollar amount invested will be higher, even though the amount invested in each region remains proportionally the same. Should large attacks occur later in the sample

[9] This occurred on October 15th, 2008 when the S&P500 lost 9.5 percent for its second biggest one-day loss ever.

period, we might (falsely) conclude that more FDI in a region leads to a higher reaction by using these dollar amounts. To counter this problem, we propose to instead to use shares of outward FDI per region. Unfortunately, regardless of this transformation, a one-step analysis will suffer from severe multicollinearity when using either dollar amounts or relative shares. For instance, when dollar amounts grow every year this implies a positive correlation between the regional outward FDI measures, whereas the correlation between FDI shares is negative by design: if one region receives a larger share, other regions will lose some of theirs. Splitting up the identification strategy in two steps avoids this problem. We choose to work with relative shares in FDI, since they are more stable over time than the dollar amounts, acknowledging that they still exhibit some variability over time. For example 12 percent of the investments made by the Telecom sector in 1998 were located in Asia and the Pacific, while this increased to 26 percent by 2010. Investments made by the Energy sector in Western Europe were 46 percent of the total FDI in 1998, but fell to 22 percent by 2010.

4.4.2 Second step: nonlinear reactions to terrorist attacks

To assess the impact of terrorist attacks on jump probabilities and jump sizes, we start by estimating:

$$P(n_t \geq 1|\Phi_t) = \alpha + \sum_{k=1}^{6} \beta_{1,k} TER_{k,t} + \sum_{k=1}^{6} \beta_{2,k} TER_{k,t}^2 + \gamma \text{Month} + \tau \text{DoW} + v_t \tag{4.10a}$$

$$E[J_t|\Phi_t] = \alpha + \sum_{k=1}^{6} \beta_{1,k} TER_{k,t} + \sum_{k=1}^{6} \beta_{2,k} TER_{k,t}^2 + \gamma \text{Month} + \tau \text{DoW} + v_t \tag{4.10b}$$

where $P(n_t \geq 1|\Phi_t)$ and $E[J_t|\Phi_t]$ are the jump probability and jump contribution respectively, $TER_{k,t}$ is the daily terrorism intensity score in each of the k defined regions and Month and DoW are controls for month and day-of-the-week effects. In order to obtain yearly estimates of the jump probability and jump size due to terrorism, we run this regression separately for each sector i in each year.

Since $P(n_t \geq 1|\Phi_t)$ is a probability and has values in the set $[0,1]$, we estimate Equation (4.10a) using a fractional logit approach (Papke and Wooldridge, 1996), obtaining:

$$E\left[P(n_t \geq 1|\Phi_t) \mid \mathbf{x}\right] = \Lambda\left(\alpha + \sum_{k=1}^{6} \beta_{1,k} TER_{k,t} + \sum_{k=1}^{6} \beta_{2,k} TER_{k,t}^{2} + \gamma \text{Month} + \tau \text{DoW}\right), \quad (4.11)$$

where Λ is the logistic function and Equation (4.11) is estimated using a Bernoulli log-likelihood function. Fitted probabilities of jumps associated with large terrorist attacks, $\eta_{i,k,t}$, will now lie in the $[0,1]$ range.

Furthermore, as $E[J_t|\Phi_t]$ is always negative, with values ranging between $(-\infty, 0]$, we estimate Equation (4.10b) using a Tobit model with a censoring from above at 0:

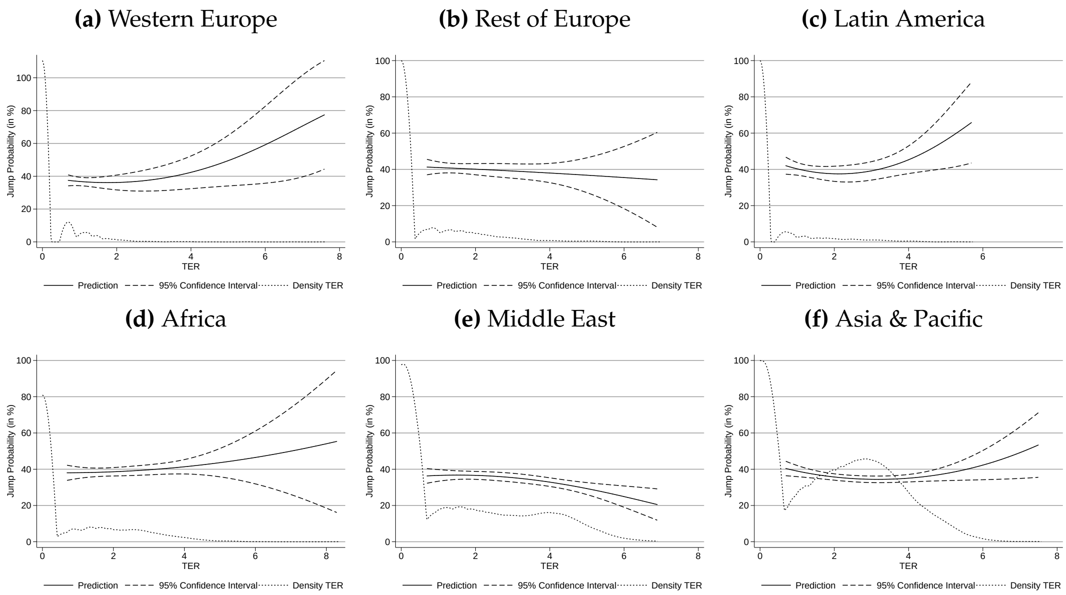
$$E[J_t|\Phi_t] = \begin{cases} E[J_t|\Phi_t]^* & \text{if } E[J_t|\Phi_t]^* < 0 \\ 0 & \text{if } E[J_t|\Phi_t]^* \geq 0 \end{cases}. \quad (4.12)$$

Fitted values for the jump size will therefore fall in the range $(-\infty, 0]$.

Of course, our objective in the end is to identify investors' reactions to events that may affect (future) cash flows. As a result, we are particularly interested in large attacks that take place in regions where a considerable FDI stock has been built up. However, what constitutes a large attack, do all regions experience those and how do investors react?

To answer these questions, we show graphically the relationship between the S&P500 jump probabilities and regional TER scores in Figure 4.2. Our aim is to find out whether there is a threshold value beyond which attacks provoke jumps. As expected, most of the patterns displayed in Figure 4.2 demonstrate a non-linear relationship, where the probability of a jump increases disproportionately with the intensity of a terrorist attack. An exception is the Middle East, where more severe attacks seem to provoke smaller reactions. At first glance this may seem odd, however given the ongoing conflict between Israel and its occupied territories, combined with the invasion and U.S. military presence in Iraq, it is possible that markets have become desensitized to attacks in this region. Figure 4.A.1 in Appendix 4.A shows the relationship for the Middle East when attacks in these countries are not used to calculate the terrorism intensity score. Compared to Figure 4.2(e), we see that the relationship between terrorism and stock market sensitivity is less downward sloping and statistically in-

Figure 4.2: S&P500 reaction to terrorism



The graphs show the relationship between the intensity of terrorism in different regions of the world and probabilities that the S&P500 experienced a jump on these attack days. Fitted values and their 95 percent confidence intervals are presented, as well as the density of the TER scores.

significant without these areas.

From Table 4.1 and Figure 4.2, we observe two things. First, the average maximum TER score per region is approximately 7.[10] Second, although the confidence intervals are quite wide due to a limited amount of large attacks, a magnitude 7 attack results in a jump probability above 50 in 4 out of the 6 regions. We therefore use this as a threshold value and condition the probability of a jump, in any given year, on the occurrence of a hypothetical attack with magnitude 7, although we use other values as a robustness check. In order to allow the reaction to this attack to vary over time, we estimate Equations (4.10a) and (4.10b) year-by-year. Given each year's and each sector's estimation, the share price reaction for each sector i to a large terrorist attacks is then:

$$\eta_{i,k,t} = E\left[(P(n_t \geq 1|\Phi_t) \mid TER_k = 7\right], \quad \text{for } t = 1, \ldots, T \tag{4.13a}$$

$$\zeta_{i,k,t} = E\left[E[J_t|\Phi_t]^* \mid E[J_t|\Phi_t] < 0 \text{ , } TER_k = 7\right], \quad \text{for } t = 1, \ldots, T \tag{4.13b}$$

where $\eta_{i,k,t}/\zeta_{i,k,t}$ is the estimated probability/size of the jump in the share price of sec-

[10]This corresponds to one large attack with 1,100 injuries and fatalities or for example 10 attacks with an average of 100 injuries and fatalities.

tor i associated with a large terrorist attack in region k in year t. Of course, in years without large attacks, the estimated relationship between terrorism and jump probabilities/sizes is expected to be relatively flat, and predicting the conditional mean given an attack outside of the observed range of actual attacks will likely yield a low jump probability/size.

As a robustness test, we will also estimate Equations (4.10a) and (4.10b) on the entire sample instead of yearly subsamples. Since every region suffers from at least one large attack (as can be seen in Figure 4.2 and Table 4.1), these estimations have an improved fit. However, this comes at a cost as we have to use average shares of FDI over the entire sample period in the next step, thereby ignoring year-to-year changes.

4.4.3 Third step: linking reactions to large attacks with FDI

Having established what constitute terrorism-effectuated jumps, the final step is to analyze to what extent the likelihood and size of these jumps are related to the U.S. FDI stock that has been built up in a region. We therefore estimate:

$$\eta_{i,k,t} = \alpha + \beta \text{FDI}_{i,k,t} + \mu_i + \tau_t + v_{i,t} \tag{4.14a}$$

$$\zeta_{i,k,t} = \alpha + \beta \text{FDI}_{i,k,t} + \mu_i + \tau_t + \nu_{i,t}, \tag{4.14b}$$

where $\eta_{i,k,t}$ and $\zeta_{i,k,t}$ are the predicted jump probability and size of sector i associated with attacks in region k occurring in year t respectively, and $\text{FDI}_{i,k,t}$ is the share of the total FDI stock of sector i that it has invested in region k in year t. Moreover, μ_i and τ_t are sector and year fixed effects respectively, and are included to control for heterogeneity between sectors and years.[11] If investors take into account the investment position of firms in each sector, we expect β to be positive for $\eta_{i,k,t}$ and negative for $\zeta_{i,k,t}$. In that case, a higher share of FDI in a region will lead to a higher probability of a jump due to a large scale attack and a more negative movement in share prices.

Since $\eta_{i,k,t}$ and $\zeta_{i,k,t}$ are predicted values of $P(n_t \geq 1|\Phi_t)$ and $E[J_t|\Phi_t]$, we again need to take into account their supports. For Equation (4.14a), we include sector fixed effects and again estimate a fractional logit.[12] Estimating Equation (4.14b) using a Tobit

[11] We do not include region fixed effects as these will absorb the cross-variation of shares between regions that we are interested in.

[12] Papke and Wooldridge (2008) caution adding fixed effects when T is small and N is large. The likelihood

approach is less straightforward, as Greene (2004) cautions that adding fixed effects biases the variance of the error term. To still account for unobserved heterogeneity due to the panel nature of the data, we estimate Equation (4.14b) using a random effects Tobit model with sector-specific random effects (see e.g. Maddala, 1987).

Given the data at our disposal, one side note is in order here. The identification strategy we follow relies on the assumption that terrorism and sector level investment are, more or less, evenly distributed across countries in the region. Since we lump together the region's investments, we allow prices to be influenced by attacks in other countries to which a sector might not be exposed. In this case, we would not expect to see a reaction on the share price as their FDI stock is not 'at stake' unless investors perceive it as a sign of regional instability. In that scenario, the potential impact would be underestimated, as we compute the average reaction to attacks in the region without being able to make the distinction whether the sector is exposed to only one particular country, or a little bit to all of them. Should we still see a reaction, even though we aggregate on a region and sector level, it would be indicative of the relative strength of these results.

4.5 Results

In this section, we retrace the steps we described above. We start by establishing whether investors react to large terrorist attacks. Then we relate these reactions to FDI stocks. Subsequently, we investigate whether investors' reaction to terrorism on foreign soil is proportional to U.S. FDI stocks in the area where the attack takes place, whether the reaction is different after 9/11, and if the information that investors receive has changed due to 9/11. We conclude this section with some robustness checks.

4.5.1 Are investors sensitive to large terrorist attacks on foreign soil?

Does a large terrorist attack on foreign soil increase the probability of a drop in U.S. share prices? And how large is the expected drop in share prices? In order to answer these questions, Table 4.3 contains the jump probabilities and jump sizes, conditional

ratio test however shows that the unobserved heterogeneity does not play a large role and the estimates are similar to a pooled version of the model. Moreover, similar to Hausman and Leonard (1997), N is fixed in our case, combined with $T = 13$ years.

Table 4.3: Average reactions to a large terrorist attack

	A. Jump probability $\eta_{i,k,t}$				B. Jump size $\zeta_{i,k,t}$			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
Western Europe	49.90	40.99	0.00	100.00	-2.33	4.05	-16.64	0.00
Rest of Europe	30.45	33.39	0.00	100.00	-0.76	1.20	-4.83	0.00
Latin America	38.77	36.13	0.00	100.00	-1.00	1.44	-7.90	0.00
Africa	28.59	22.61	0.08	89.32	-0.48	0.58	-2.97	-0.01
Middle East	30.55	24.58	0.00	96.66	-0.73	1.43	-11.60	0.00
Asia & the Pacific	32.94	22.24	0.04	89.96	-0.67	0.84	-6.68	-0.01
Contribution TER (in %)	10.91	5.69	3.10	29.78	10.22	6.19	2.30	39.81

The table displays summary statistics for the predicted jump probabilities (Panel A) and predicted jump sizes (Panel B) associated with a large attack in each of the regions. The probabilities and sizes are averaged out over year and sector. A large terrorist attack is defined as having a TER score of 7, which corresponds to 1,100 injuries and fatalities. The predictions are obtained by estimating Equations (4.11) and (4.12) on a year and sector basis. A Shapley decomposition of McFadden's pseudo-R^2 is performed to determine the contribution of the combined TER terms .

on a large attack, as defined by Equations (4.11) and (4.12).

We observe that, during the entire sample period, the region where large attacks lead to the highest average predicted jump probability and jump size is Western Europe. On average, the probability that at least one jump occurred in one of the sectors due to a large terrorist attack is 49.90 percent, and the average size of that jump is -2.33 percent. Other regions where attacks lead to noticeable reactions on U.S. stock markets are Latin America and Asia and the Pacific, where the average $\eta_{i,k,t}$ are 38.77 percent and 32.94 percent respectively. The average predicted jump sizes are -1.00 percent for Latin America and -0.76 percent for Asia and the Pacific respectively, and are markedly lower than for Western Europe.

To analyze how much predictive power the TER terms add, we perform a Shapley decomposition, which is based on the pseudo-R^2 of each regression, and is shown in Table 4.3. On average, around 10% of the predictive power in the regression comes from the combined TER terms, although for certain year/sector combinations this is as high as 29% for $\eta_{i,t,k}$ and 39% for $\zeta_{i,t,k}$.

Of course, since large terrorist attacks are still rare, share prices are not always expected to jump. For instance, in 2006 our model predicts a jump probability for Western Europe's Industrials sector equal to 0 percent, whereas in 2009 the jump probability for the same sector was 12.5 percent, although the actual terrorist activity in the region in those years was low and quite similar.[13]

4.5.2 Is investors' reaction to large attacks related to the FDI stock that is 'at stake'?

Now that we have established investors' average reaction to large terrorist attacks, the next question is whether this reaction depends on the FDI stock built up by U.S. firms in the region where a large attack takes places. To answer that question, we regress the jump probabilities and jump sizes on FDI stocks. We expect that jump probabilities increase with larger FDI stocks, and we expect to see larger *negative* jumps as FDI stocks increase.

Indeed, this is what we observe from Table 4.4. The coefficients on the share of received FDI are significant and as expected: higher shares of outward U.S. FDI in a region lead to higher jump probabilities and larger negative jump sizes in response to high TER attacks.[14]

Since the interpretation of the magnitude of the coefficients in Table 4.4 is not straightforward due to the nonlinear nature of the models, we plot the relationship between FDI stocks and jump probabilities/sizes in Figure 4.3. We observe that the probability of a jump increases from the unconditional expected value of 30 percent when no investments take place, to 70 percent when a region receives all sector level FDI. The size of the jump due to large terrorist attacks also increases with received FDI, leading to a stock market reaction of -3 percent if a region receives all FDI from that sector. The highest share a region received is found in 2006 when the Telecom sector had 70 percent of its FDI stock invested in Western Europe. Using these estimates, a large attack in that year would have led to a drop of -2.5 percent in its share price, with a 60 percent probability of a jump.

These results show that the presence of U.S. firms in regions suffering from terror-

[13]The highest daily TER scores for Western Europe in 2006 and 2009 are 2.64 and 2.48, respectively.

[14]Year and sector fixed effects are included in the fractional logit case, and only year controls in the random effects Tobit case. The sector fixed effects are found to be jointly insignificant in the fractional logit estimation, meaning that the estimates are similar to a pooled version of the model. For the random effects Tobit model, the unobserved heterogeneity in $\zeta_{i,k,t}$ does not play a large role as the random sector effect contributes only 9 percent of total variance. However, a likelihood ratio test comparing the random effects with a pooled version shows that the random effect is significantly different from zero.

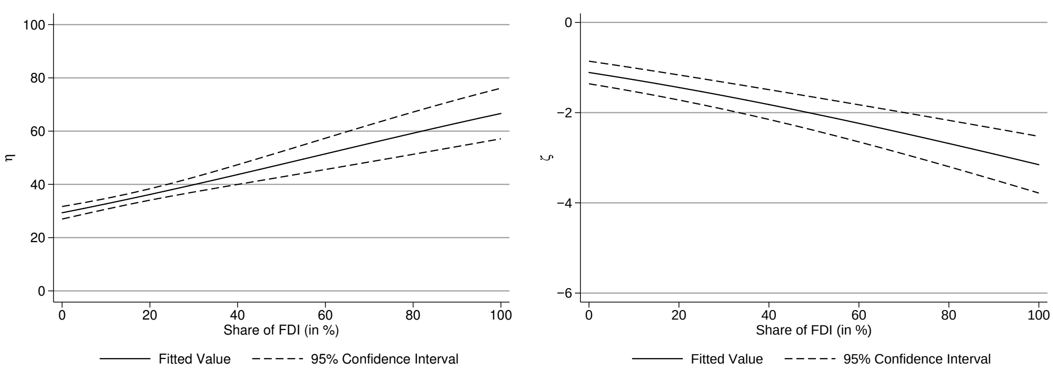
Table 4.4: Share price reactions and FDI stock

	Jump probability $\eta_{i,k,t}$	Jump size $\zeta_{i,k,t}$
α	-0.917***	-1.235***
	(0.229)	(0.307)
$\text{FDI}_{i,k,t}$	1.723***	-2.539***
	(0.270)	(0.331)
Year FE	Included	Included
Sector FE	Included	
N	696	696
L	-325.27	-1380.07
ρ		0.09
LR Test 1		48.88***
LR Test 2	1.59	

The table shows panel regression results for Equations (4.14a) and (4.14b), where reactions of sectors to large terrorist attacks in regions are regressed on the share of FDI they receive. $\eta_{i,k,t}$ is estimated using fractional logit, whereas $\zeta_{i,k,t}$ is estimated using a random effects Tobit model. Standard errors are reported in parentheses, and are robust for $\eta_{i,k,t}$. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent. L is the log likelihood value, ρ is the fraction of variance due to the unobserved heterogeneity μ_i and is defined as $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\epsilon^2}$. LR Test 1 is a likelihood ratio test of $\sigma_\mu = 0$ in $\zeta_{i,k,t}$. LR Test 2 is a likelihood ratio test for the joint significance of the sector fixed effects in $\eta_{i,k,t}$.

Figure 4.3: Are share price reactions proportional to FDI stock?

(a) Jump probability $\eta_{i,k,t}$ **(b)** Jump size $\zeta_{i,k,t}$



The graphs show the jump probability and size in reaction to a large terrorist attack in a region, depending on the share of FDI this region receives. Fitted values and their 95 percent confidence intervals are shown over the response surface for the regressions in Equations (4.14a) and (4.14b), and Table 4.4.

Table 4.5: The impact of 9/11 on investors' reaction

	A. Jump probability $\eta_{i,k,t}$			B. Jump size $\zeta_{i,k,t}$		
	Before 9/11	After 9/11		Before 9/11	After 9/11	
	1998 - 2001	2002 - 2006	2007 - 2010	1998 - 2001	2002 - 2006	2007 - 2010
mean	0.336	0.306	0.427	-0.742	-0.816	-1.489
(std. dev.)	(0.273)	(0.306)	(0.354)	(1.374)	(2.072)	(2.467)
α	-0.966***	-1.625***	-1.296***	-0.666***	0.120	-0.867**
	(0.294)	(0.343)	(0.301)	(0.257)	(0.276)	(0.411)
$\text{FDI}_{i,k,t}$	0.537	1.439***	3.532***	-0.356	-2.449***	-4.745***
	(0.443)	(0.428)	(0.536)	(0.428)	(0.526)	(0.688)
Year FE	Included	Included	Included	Included	Included	Included
Sector FE	Included	Included	Included			
N	210	270	216	210	270	216
L	-99.31	-118.03	-99.77	-348.15	-536.27	-460.52
ρ				0.17	0.04	0.15
LR Test 1				22.51**	3.10***	20.56***
LR Test 2	5.41	1.37	3.40			

The table shows panel regression results of Equations (4.14a) and (4.14b) for different subsamples, where reactions of sectors to large terrorist attacks in regions are regressed on the share of FDI they receive. $\eta_{i,k,t}$ is estimated using fractional logit, whereas $\zeta_{i,k,t}$ is estimated using a random effects Tobit model. Standard errors are reported in parentheses, and are robust for $\eta_{i,k,t}$. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent. L is the log likelihood value, ρ is the fraction of variance due to the unobserved heterogeneity μ_i and is defined as $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\epsilon^2}$. LR Test 1 is a likelihood ratio test of $\sigma_\mu = 0$ in $\zeta_{i,k,t}$. LR Test 2 is a likelihood ratio test for the joint significance of the sector fixed effects in $\eta_{i,k,t}$.

ism directly affects their share prices. The higher the proportion of investments made in a region, the more the share prices of these sectors react to large terrorist attacks taking place there. This suggests that investors do on average take into account the investment positions of firms belonging to a sector and seem to adjust expectations on cash flows when they are dealt with the negative exposure of terrorism.

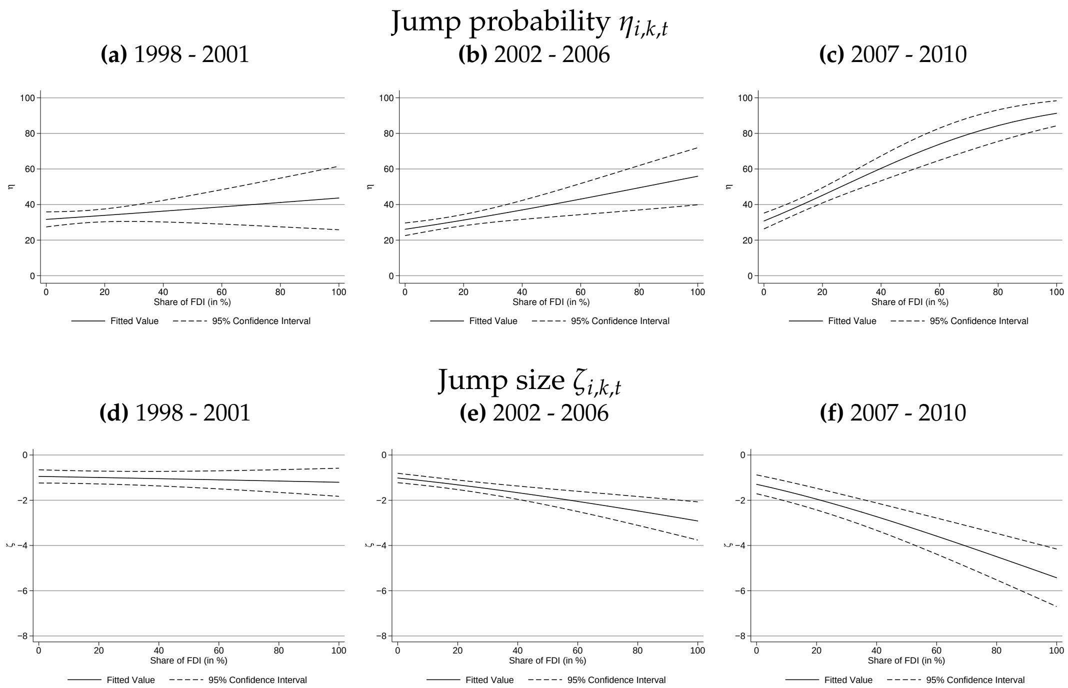
4.5.3 Has 9/11 made investors more sensitive to attacks abroad?

Since 9/11 occurs in the middle of our sample, the question arises to what extent our results so far reflect the post 9/11 state of the world. In order to find out, we split our sample and compare the relationship between investors' reaction and U.S. FDI stock before and after 9/11, as well during the recent financial crisis.

Table 4.5 shows the results for each of these periods, relating jump probabilities and jump sizes conditional on large terrorist attacks to FDI stocks.[15] We observe that FDI stocks had no relation to jumps prior to 9/11. After 9/11, higher FDI stocks resulted in higher jump probabilities and more negative jumps.

[15]The jump probability estimates are, however, still based on the entire sample.

Figure 4.4: Changes in sensitivity of investors after 9/11



The graphs show for different subsamples the jump probability and size in reaction to a large terrorist attack in a region, depending on the share of FDI this region receives. Fitted values and their 95 percent confidence intervals are shown over the response surface for the regressions in Equations (4.14a) and (4.14b) and Table 4.5.

Figure 4.4 plots the predictive margins, as the size of the coefficients can not be properly interpreted due to the nonlinear nature of both models. We indeed see that in the first subsample, investors were less wary of large attacks occurring in regions receiving proportionally more FDI. The confidence intervals show that the effects are not statistically significant. Between 2002 and 2006, the slope shifts and we see a gradual increase in stock market reaction to attacks in high FDI regions. The role of FDI becomes even more apparent in the last subsample. In both periods after 9/11, the relationship is statistically significant. In the second subsample, a sector investing all of its FDI in a region under attack experiences on average a jump of -3 percent with a probability of 60 percent. In the last subsample, this same sector would see a near-certain jump of -5.5 percent in its share price.

4.5.4 Robustness

In the previous section we have provided evidence that a terrorist attack at home can act as a wake up call to investors, making them aware of attacks elsewhere and thereby changing the way they evaluate the frequency and probability of terrorist attacks. Moreover, we have shown that the reaction of investors is related to the FDI stock 'at stake'. Here we examine the robustness of these results.

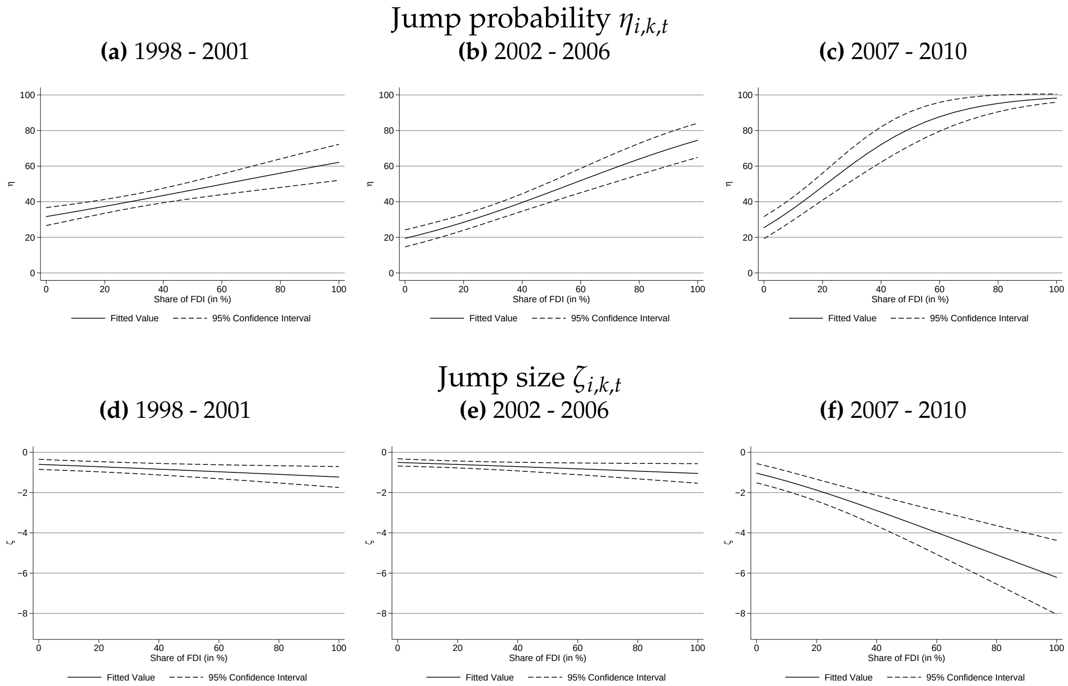
From Figure 4.4 we have seen that the relationship between stock market jumps and large terrorist attacks in regions receiving more U.S. FDI has become economically and statistically significant after 9/11. Following attacks in Western Europe, the largest receiver of U.S. FDI, we find that the relationship intensifies even more and is strongest during the last subsample. The fact that large jumps were more prevalent during the financial crisis is controlled for using time fixed effects. Even so, estimating the last subsample without the turbulent year 2008 does not change our results.[16]

Another question is to what extent our findings in Figure 4.4 are the result of the fact that we estimate Equations (4.10a) and (4.10b) for each year, rather than for our entire sample. In order to find out, we re-estimate them using the entire sample while adding year fixed effects. As we already concluded from Table 4.1, all regions have seen at least one large attack and by using these estimations we therefore avoid having to predict the average reaction to large attacks when they do not occur in a given year. The downside of this strategy is that we are forced to use average shares of outward FDI, disregarding their yearly variability. In Appendix 4.A, Table 4.A.3 shows the estimation results, while Figure 4.A.2 displays the impact of an increase in FDI stock on the jump probabilities and jump sizes for the whole sample. In Figure 4.5 we show the pooled estimation on each of the subsamples. The results are in line with what we have found so far. In fact, if anything, we find that although the average jumps are somewhat smaller, the changes in jump probabilities after 9/11 are more remarkable, especially in the last subsample. It appears that the impact of 9/11 has indeed lasted for a long time.

How relevant is the size of a terrorist attack? Is it the case that the large attacks are what drives investors' reactions? To find out, we check whether conditioning the

[16] See Table 4.A.2 in Appendix 4.A for the estimations. The coefficients are actually larger in size when estimating without the year 2008.

Figure 4.5: Reaction to terrorism and share of FDI - Pooled estimation - Split sample



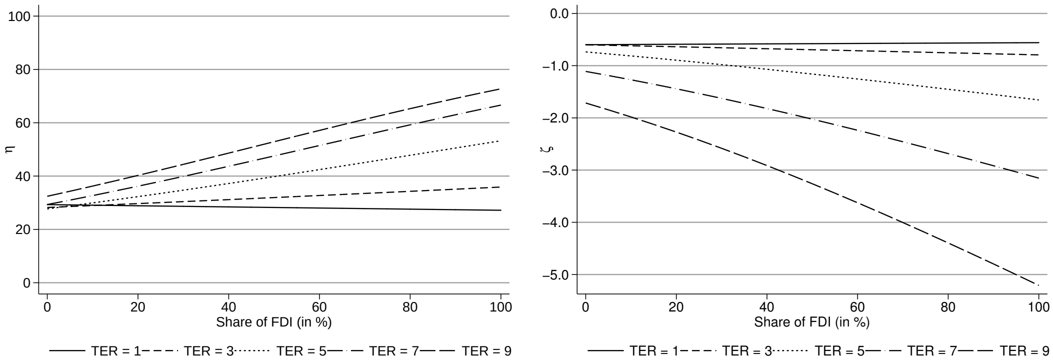
The graphs show for a pooled version of the model, the jump probability and size in reaction to a large terrorist attack in a region during different subsamples, depending on the share of FDI this region receives in this subsample. Fitted values and their 95 percent confidence intervals are shown over the response surface for the regressions in Equations 4.14a and 4.14b, and Table 4.A.3.

yearly jump forecasts on smaller attacks changes our conclusion. Table 4.A.4 in Appendix 4.A shows the regression results for levels of $TER = 1$, $TER = 3$, $TER = 5$, baseline specification $TER = 7$ and $TER = 9$, while Figure 4.6 plots the jump probabilities and jump sizes for increasingly large attacks. When terrorist attacks are of a smaller magnitude, the sector indices do not move and the reaction is very small, even if a sector would make its investments in one single region. As terrorist attacks become increasingly large, the share prices react more to attacks in regions receiving more FDI. For extreme terrorism of $TER = 9$, sectors investing all of their FDI in one region would experience a jump of -5.2 percent with a probability of 72 percent. The exercise shows that this relationship only exists for the largest of observed attacks. In Table 4.1 we saw that the mean TER scores lie between 0.23 and 1.74. If we had estimated using these mean values, we would have been unable to see the reaction both in terms of jump probability and in jump size.

Finally, we check whether the results are driven by the large attacks in London (2004) and Madrid (2005), as Western Europe received the bulk of U.S. outward FDI.

Figure 4.6: Reaction to terrorism and share of FDI - Variable TER

(a) Predictive Margins η 1998 - 2010 **(b)** Predictive Margins ζ 1998 - 2010



The graphs show the jump probability and size in reaction to different-sized attacks in each region, depending on the share of FDI this region receives. Fitted values over the response surface of the regressions in Equations (4.14a) and (4.14b), and Table 4.A.4 are shown.

Similarly, a country like China receives a lot of FDI, but there are only 21 attack days recorded in the GTD during the sample period. We ran the analysis by excluding the London and Madrid attacks, as well as a separate analysis without China. When excluding these observations, the results are qualitatively similar. In both cases the relationship between FDI and the reaction to large terrorist attacks is negative and statistically significant.[17]

4.5.5 What can explain the change in investor behavior?

The analyses show that investors appear to place more emphasis on terrorist attacks on foreign soil after 9/11, but it remains unclear what can explain this change. The world has changed in more ways than one following 9/11. First, Straetmans et al. (2008) find that stock markets have structurally changed after the 9/11 attacks, and point out that this might be caused by the perceived risk of new attacks. Second, in the period after 9/11, the U.S. commenced the War-on-Terror and invaded Afghanistan and Iraq, as a response to which terrorists attacked public transportation in Madrid (2004) and London (2005). These experiences showed that Western Europe, beneficiary of the bulk of U.S. FDI, could also be subject to terrorist attacks. Finally, with the developed world suffering from the declining U.S. housing market, more focus was placed on emerging markets as a source of growth and profit. However, while

[17] Results are available on request.

investors and companies shifted more of their investments to regions like Asia and the Middle East, these regions also saw the bulk of terrorist attacks (see Table 4.1).[18]

Below, we develop and discuss several explanations for this change in investor behavior. We first examine possible rational responses by investors, before positing a behavioral counterpart.

Terrorism magnitude A first possibility is that acts of terrorism have occurred more frequently after 9/11, or that they have increased in severity, thereby decreasing future cash flows more compared to the pre-9/11 attacks. To test this possibility, we plot the distribution of the terrorism index for each of the regions, before and after 9/11, in Figure 4.7. Overall, the distributions have not changed significantly since 9/11. The average value for the index is in fact *lower* in Western Europe, the Rest of Europe, Latin America and Africa after 9/11. On the other hand, the average value of the index is somewhat higher in the Middle East and Asia and the Pacific as there were more and larger attacks. However, extreme attacks occur in all regions both before and after 9/11. A similar pattern is visible when we only use U.S. casualties or injuries to calculate the TER scores.[19] Furthermore, both before and after 9/11, over 90% of the attacks are classified as having only minor property damage for those limited number of attacks where this data is available.[20]

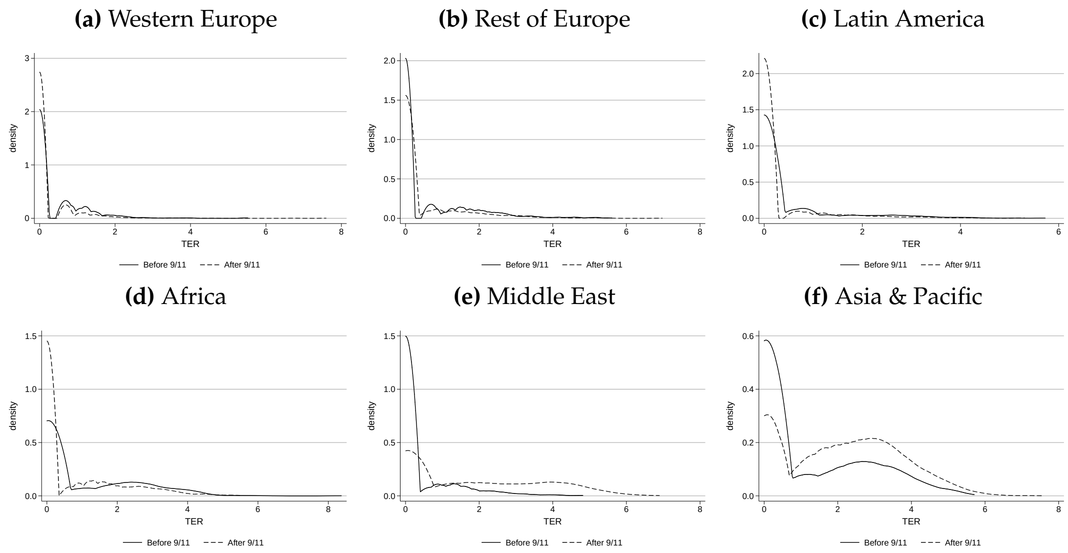
Terrorism risk perception A second possibility is that the ex-ante terrorism risk perception is higher after the 9/11 attacks, as alluded to by e.g. Straetmans et al. (2008), or documented by Bozzoli and Müller (2011) using the London bombings. It is possible that the 9/11 attacks have lead to an increase in the anticipation of new attacks and a higher risk perception, which can have important implications for asset prices (see e.g. Wachter, 2013). While we are not able to measure this risk perception directly, Drakos and Müller (2014) use survey data to show that risk perception is mainly driven by terrorism activity. For the purpose of our analysis, we have seen that this terrorism activity has not changed dramatically after 9/11, which in concordance with Drakos

[18] The United Nations Conference on Trade and Development (UNCTAD) (2011) reports that in 2011 52 percent of the world FDI inflows occurred in Developing and Transition Economies, up from 33 percent in 2007. Based on the UNCTAD database on FDI stock, the share of these markets in the total world FDI stock increased from 29 percent in 2007 to 35 percent in 2010.

[19] Results are available upon request.

[20] Minor damage is defined as likely below $1.000.000. See LaFree and Dugan (2007) for more details.

Figure 4.7: Terrorism before and after 9/11

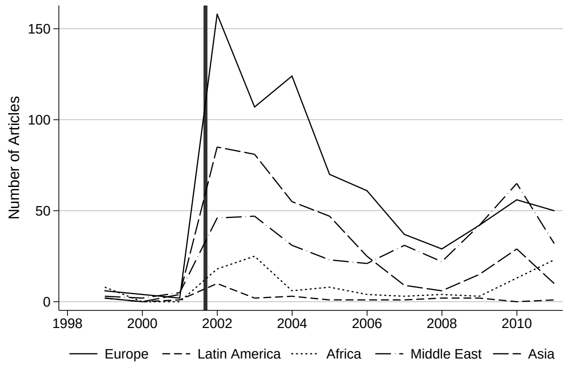


The graphs show the distribution (as a kernel density plot) of the terrorism index, before and after the 9/11 attacks. Kolmogorov-Smirnov tests for differences in the distribution among days on which attacks take place indicate that only for Africa and the Middle East there is a change after 9/11.

and Müller (2014) suggests that an increase in risk perception would not have played a role throughout the entire post-9/11 part of our sample. Moreover, even if the risk perception had increased after the 9/11 attacks, it is unlikely that it would have remained as high throughout the remaining part of the sample during which our results are the most significant.

Media A third possible explanation is that since the 9/11 attacks media coverage of other acts of terrorism has increased, or has given investors more accurate information regarding the damages that have taken place. For instance, Sunstein (2003) warns for exaggerated risk perceptions due to the media, and (Melnick and Eldor, 2010) find that the economic damage caused by terrorist attacks increases with the amount of reporting. Media coverage can therefore be an explanation why the change in investor behavior documented in the previous section. While, unfortunately, we can not measure the informational content of the news provided, we can measure the amount of reporting on terrorism during our sample. In Figure 4.8, we proxy the amount of reporting in the U.S. about terrorism on foreign soil by plotting the yearly number of articles that appeared in the Wall Street Journal for searches on 'terrorism' or 'terrorist'

Figure 4.8: Reporting about terrorism on foreign soil



This graph shows the number of Wall Street Journal articles per year for a Boolean search on '(terrorism OR terrorist) AND X', where X is each of the regions. The search is conducted using the Lexis-Nexis newspaper database. Since the region 'Rest of Europe' is not properly defined outside the context of this study, we group it together with 'Western Europe' in one large 'Europe' region.

in each of the regions identified in our analysis. As expected, we see a coverage spike following the 9/11 attacks, driven mainly by articles relating to terrorism and the Europe, Asia and Middle East regions. Prior to 9/11 (1998-2000), there are an average of 3 articles a year on terrorism in these three regions, increasing to 65 articles in the 5 years following the attacks. However, the overall trend is downwards and, with the exception of the Middle East, seems to be leveling out around 2006. These numbers have to be interpreted with some care: the number of articles depend positively on the occurrence of large attacks, and not every article could give investors accurate information on the probability of a terrorist attack occurring. Nonetheless, based on these figures, we infer that despite the media coverage of terrorism being lower compared to the years following 9/11, the relationship between price jumps, terrorism and FDI stocks remained significant and became stronger. This is also in line with Melnick and Eldor (2010), who show that the economic impact of terrorist attacks attributable to media coverage diminishes over time, indicating that the news-value for investors at the end of the sample period is potentially lower than right after the attacks.

Terrorism Risk Insurance Act A fourth possibility is that it has become more costly for U.S. firms to protect or insurance themselves against losses due to terrorism. For instance, Brown et al. (2004) document a negative announcement effect of the events leading up to the Terrorism Risk Insurance Act of 2002 (TRIA). They explain their finding by reduced market expectations of federal assistance following new acts of terrorism, as the Act obligated insurers to offer insurance specifically covering acts of terrorism. If investors believe that attacks abroad will increase insurance premia for the companies involved, share prices could drop when new attacks take place. However, it seems that the cost of terrorism insurance has come down by over 55% since 2004: in 2004 the cost to insure one million dollars was $57 (0.0057%), this decreased to $25 per million (0.0025%) in 2009 (Marsh, 2005, 2010). Moreover, this type of insurance has become more popular to the extent that recent studies estimate around 60% of firms have signed up. One possible reason is that firms might want to avoid reputation costs (Marsh, 2010; Michel-Kerjan et al., 2011), something that could be especially true for the companies we consider. Finally, while the cost of insurance might decrease profitability, it also theoretically covers the downside when terrorist attacks do occur. In that case, we should not be able to find the stock market reaction to large terrorist attacks which we have documented.

Availability heuristic The 'rational' explanations above seem inadequate to fully explain the change in investor behavior we have documented. The question what could have caused this shift is still unanswered, making a 'behavioral' explanation more plausible. The evidence we have presented seem consistent with the possibility that 9/11 has served as a wake-up call, creating awareness among investors about the potential impact of terrorist attacks. People tend to have quite visceral reactions to acts of terror compared to other extreme events. For instance, Kip Viscusi (2009) points out that individuals value prevention of terrorism more than prevention of natural disasters or traffic deaths, even though the latter two lead to more fatalities per year. Moreover, Johnson et al. (1993) documents that individuals valued insurance against acts of terror higher than insurance covering all causes (including terrorism). Since the 9/11 attacks hit the U.S. right in the financial heart, it is possible that their behavior afterwards is influenced by the availability heuristic (Tversky and Kahneman, 1973), probability neglect (Sunstein, 2003) and local thinking (Gennaioli and Shleifer, 2010). The availability heuristic states that events which are more salient, such as terrorism, are more easily remembered and are judged to be more common than they in reality are. This in turn can lead to biased expectations that an attack will occur. Probability neglect can also lead to biased expectations, as individuals focus only on the possible bad outcome, without taking into account it is highly unlikely to materialize. Given the evidence presented here, it is possible that when news of terrorist attacks abroad reaches investors, they remember 9/11 and its (economic) consequences, leading them to sell their shares of firms that are possibly exposed to this attack. This explanation can give the appearance that after 9/11 investors overreact to terrorism. However, Figure 4.4 shows that the stock market reaction to large attacks is proportional to the share of investment after 9/11, but not before. It seems, therefore, that 9/11 has made it easier for U.S. investors to bring to mind terrorist attacks, including those that take place on foreign soil. Instead of overreacting in the post-9/11 period, investors appear to have under-reacted in the period prior to 9/11, with 9/11 serving as a wake-up call acknowledging the dangers of terrorism, even for attacks occurring abroad. This explanation is in line with the experience hypothesis formulated by (Malmendier and Nagel, 2011), who show that extreme events can have a longlasting effect on the behavior of households and investors. Given the discussion above, we consider this explanation the most likely.

4.6 Conclusion

We have examined how terrorism in different geopolitical regions of the world has spilled over to U.S. financial markets through the foreign presence of U.S. firms, and how this relationship has changed after 9/11. We document that share prices react negatively to large terrorist attacks on foreign soil, and that this reaction is related to the FDI stock of U.S. firms on that soil. However, in order for investors to act this way, the 'message' unfortunately has had to hit home first: the relationship is only significant, both statistically and economically, after the tragic events of September 11, 2001, indicating disaster myopia consistent with the availability heuristic. However, given that the frequency of attacks has not changed materially after 9/11, the relationship between share prices, FDI and terrorist attacks abroad has stayed strong even as the media coverage of these attacks has come down from its peak levels post-9/11. Presented with this evidence, we conclude that 9/11 was the experience that created awareness of the potential impact of terrorist attacks among investors, and, in line with the experience hypothesis, explains the change in their behavior afterwards.

The results in this chapter are in line with a growing literature (see e.g. Abadie and Gardeazabal, 2003; Drakos, 2010a) that finds that the (global) activity of firms and/or sectors leads to sensitivity in their share prices as a reaction to acts of terrorism. In an increasingly globalized world, this has an impact on both companies and investors. On the one hand, investors need to take into account the geopolitical situation in regions where firms locate their FDI before they invest in this company. On the other hand, multinationals valuing the stability of their share price also need to take this into account before investing in these regions. The documented relationship between foreign presence of firms, terrorism and their share price is likely to become even more important in the coming years: U.S. firms have increasingly built up their presence in Asia and the Middle East, yet these regions have seen the bulk of terrorist attacks since 1998, and are likely to continue to pose a geopolitical risk in the near future. A better understanding of sensitivity to terrorism, preferably using firm level investment positions, is therefore a promising avenue for future research.

Appendix 4.A Additional figures and tables

Estimates of the GARCH-Jump model for the S&P500 and sector indices are shown in Table 4.A.1. The results show that the estimated average return, μ, is not significantly different from zero and is in line with the actual average return in the sample period. The main coefficients of interest are those that govern the jump dynamics. The autoregressive coefficient in the jump intensity equation, ρ, is close to 1 for all indices, indicating that the jump intensity and jump probabilities advance smoothly over time. It also indicates that, like volatility, jumps exhibit clustering. A likelihood ratio test, testing if $\rho = \gamma = 0$, shows that the null hypothesis that the jump intensity is time-invariant can be rejected.[21] The average jump size, θ, is negative and significant for all indices ranging between -0.4 percent (Telecom) and -1.7 percent (Energy). Since we multiply the expected number of jumps ($E[n_t|\Phi_t]$) with the average jump size (θ) to obtain the expected jump contributions ($E[J_t|\Phi_t]$), this means that the expected jump contribution is always negative although the realized jump contribution can be positive. Similar to Maheu and McCurdy, we find that for some indices the impact of jumps on the conditional mean tends to be centered around zero (more specifically in Consumer Staples, Health Care, Materials and Telecom as indicated by the values of θ and δ). The authors however show that even in the case that $\theta = 0$, the jump dynamics can still lead to tail realizations.

The unconditional expected level of jumps, $E\lambda_t$, shows that jumps are more likely to occur during our sample period compared to the more stable indices chosen by Maheu and McCurdy. The difference is due to higher estimates of autoregression in the jump intensity, ρ, and the jump intensity constant, λ_0. One of the reasons we find a higher expected jump intensity could be that Maheu and McCurdy use an estimation window between 15 and 40 years up to the end of 2001, while our estimation period is thirteen years in which there were three distinct crises (the 1998 crisis, the crash of the internet bubble and the recent financial crisis). Finally, Maheu and McCurdy suggest that the effect of jumps on returns is best measured by the unconditional variance of jump innovations, which is reported in the last row of Table 4.A.1. This average variance due to jumps is highest for the Financials, Energy and IT indices.

[21] The constraint is similar to $\lambda_t = \lambda$.

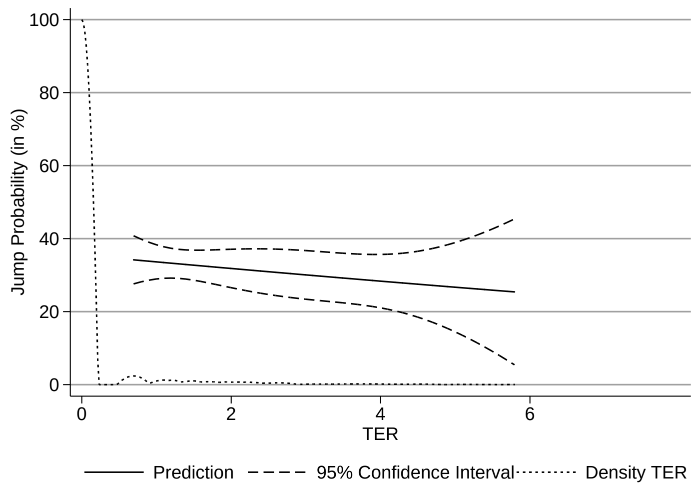
Table 4.A.1: GARCH-Jump estimates

$$r_t = \mu + \phi r_{t-1} + \epsilon_{1,t} + \epsilon_{2,t}, \quad \epsilon_{1,t} = \sigma_t z_t, \quad z_t \sim \text{NID}(0,1),$$
$$\epsilon_{2,t} = \sum_{k=1}^{n_t} Y_{t,k} - \theta\lambda_t, \quad Y_{t,k} \sim N(\theta, \delta^2), \quad \lambda_t = \lambda_0 + \rho\lambda_{t-1} + \gamma\xi_{t-1},$$
$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2, \quad \epsilon_{t-1} = \epsilon_{1,t-1} + \epsilon_{2,t-1}.$$

	S&P500	C. Discretionary	C. Staples	Energy	Financials	Health Care	Industrials	IT	Materials	Telecom	Utilities
μ	0.005	0.004	0.016	0.034	0.001	-0.007	0.016	-0.021	0.033	0.015	0.006
	(0.016)	(0.018)	(0.013)	(0.024)	(0.018)	(0.017)	(0.017)	(0.024)	(0.023)	(0.020)	(0.017)
ϕ	-0.044**	0.032*	-0.067***	-0.033*	-0.032*	-0.010	-0.010	-0.003	0.000	-0.032*	0.028
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.019)	(0.018)	(0.019)
ω	0.004***	0.004**	0.003***	0.017***	0.007***	0.005**	0.006***	0.009***	0.008***	0.009***	0.016***
	(0.001)	(0.002)	(0.001)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
α	0.014***	0.022***	0.018***	0.020***	0.035***	0.014**	0.014***	0.018***	0.016***	0.016***	0.045***
	(0.003)	(0.005)	(0.005)	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)	(0.005)	(0.006)	(0.009)
β	0.968***	0.962***	0.968***	0.964***	0.949***	0.969***	0.971***	0.969***	0.972***	0.969***	0.922***
	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.011)	(0.007)	(0.005)	(0.007)	(0.008)	(0.013)
λ_0	0.020***	0.019***	0.020***	0.011***	0.011***	0.027***	0.013***	0.017***	0.019**	0.009**	0.015***
	(0.005)	(0.005)	(0.006)	(0.004)	(0.003)	(0.007)	(0.004)	(0.006)	(0.008)	(0.004)	(0.005)
ρ	0.974***	0.974***	0.957***	0.970***	0.984***	0.963***	0.975***	0.976***	0.957***	0.978***	0.917***
	(0.005)	(0.007)	(0.012)	(0.008)	(0.004)	(0.010)	(0.005)	(0.006)	(0.014)	(0.008)	(0.021)
γ	0.417***	0.468***	0.752***	0.455***	0.541***	0.801***	0.410***	0.564***	0.821***	0.578***	0.485***
	(0.044)	(0.067)	(0.110)	(0.066)	(0.090)	(0.116)	(0.060)	(0.073)	(0.170)	(0.145)	(0.107)
θ	-1.051***	-1.149***	-0.421***	-1.700***	-1.256***	-0.470***	-1.290***	-1.564***	-0.636***	-0.402***	-1.572***
	(0.098)	(0.115)	(0.076)	(0.192)	(0.169)	(0.088)	(0.137)	(0.187)	(0.197)	(0.148)	(0.250)
δ	0.187	0.084	0.755***	0.007	0.368	0.943***	0.241	0.016	1.161***	1.387***	0.146
	(0.163)	(0.420)	(0.076)	(0.582)	(0.244)	(0.077)	(0.217)	(0.257)	(0.148)	(0.205)	(0.708)
L	-4,833.99	-5,373.82	-4,123.81	-5,980.20	-5,874.87	-4,860.22	-5,215.49	-6,238.37	-5,830.54	-5,580.58	-4,869.79
N	3,270	3,270	3,270	3,270	3,270	3,270	3,270	3,270	3,270	3,270	3,270
LR Test	153.99***	94.96***	55.10***	72.72***	79.84***	55.87***	114.88***	68.04***	54.88***	26.93***	56.41***
$E\lambda_t$	0.769	0.731	0.465	0.367	0.687	0.730	0.520	0.708	0.442	0.409	0.181
$E\epsilon_{2t}^2$	0.891	1.001	0.352	1.071	1.183	0.820	0.919	1.703	0.763	0.816	0.440

The table shows the regression results of the GARCH-Jump model. Standard errors are reported in parentheses. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent. The dependent variables are daily returns in percentages, where a 1 percent return is denoted as 1. L is the log likelihood value. LR Test is the Likelihood Ratio test of $\rho = \gamma = 0$, testing for time-varying jump intensity. $E\lambda_t$ is the implied unconditional jump intensity and is calculated as $\frac{\lambda_0}{1-\rho}$. $E\epsilon_{2,t}^2$ is the unconditional variance of jump innovations and is equal to $\frac{(\theta^2+\delta^2)\lambda_0}{1-\rho}$.

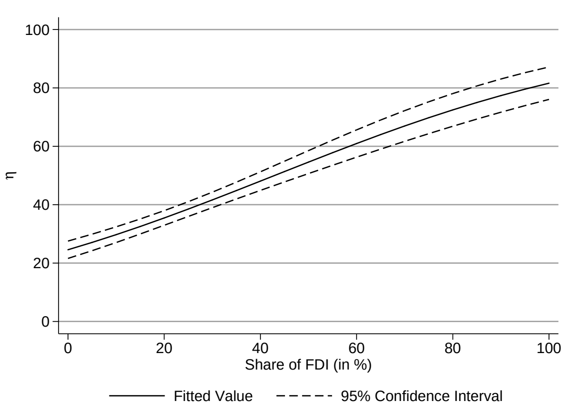
Figure 4.A.1: S&P500 reaction to terrorism in the Middle East - Excluding attacks in Iraq, Israel and Occupied Territories



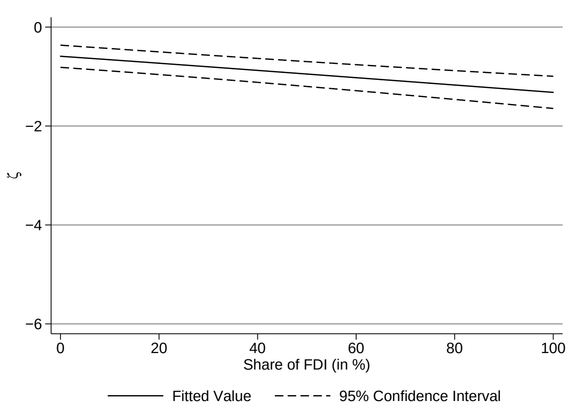
The graph shows the relationship between the intensity of terrorism in the Middle-East – excluding Iraq, Israel and the occupied territories – and probabilities that the S&P500 experiences a jump on these attack days. Fitted values and their 95 percent confidence intervals are shown for a univariate quadratic regression of the S&P500 jump probabilities on the TER score for days where attacks took place in this region.

Figure 4.A.2: Reaction to terrorism and share of FDI - Pooled estimation

(a) Predictive Margins η



(b) Predictive Margins ζ



The graphs show for a pooled version of the model, the jump probability and size in reaction to a large terrorist attack in a region, depending on the share of FDI this region receives. Fitted values and their 95 percent confidence intervals are shown over the response surface for the regressions in Equations (4.14a) and (4.14b) and Table 4.A.3.

Table 4.A.2: Investors' reaction conditional on FDI Stock in 2006-2010, excluding 2008

	Jump probability $\eta_{i,k,t}$	Jump size $\zeta_{i,k,t}$
mean	0.365	-1.159
(std. dev.)	(0.349)	(2.363)
α	-1.657***	-0.829**
	(0.353)	(0.358)
$\text{FDI}_{i,k,t}$	4.120***	-4.973***
	(0.608)	(0.816)
N	162	162
L	-72.20	-345.30
ρ		0.06
LR Test 1		3.53***
LR Test 2	1.45	

The table shows the panel regression of Equations (4.14a) and (4.14b) without 2008. Reactions of sectors to large terrorist attacks in regions are regressed on the share of FDI they receive. $\eta_{i,k,t}$ is estimated using fractional logit, whereas $\zeta_{i,k,t}$ is estimated using a random effects Tobit model. Standard errors are reported in parentheses, and are robust for $\eta_{i,k,t}$. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent. L is the log likelihood value, ρ is the fraction of variance due to the unobserved heterogeneity μ_i and is defined as $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\epsilon^2}$. LR Test 1 is a likelihood ratio test of $\sigma_\mu = 0$ in $\zeta_{i,k,t}$. LR Test 2 is a likelihood ratio test for the joint significance of the sector fixed effects in $\eta_{i,k,t}$.

Table 4.A.3: Regression output - Pooled estimation

	A. Jump probability $\bar{\eta}_{i,k}$				B. Jump size $\bar{\zeta}_{i,k}$			
	1998 - 2010	1998 - 2001	2002 - 2006	2007 - 2010	1998 - 2010	1998 - 2001	2002 - 2006	2007 - 2010
mean	0.342	0.365	0.277	0.438	-0.699	-0.662	-0.557	-1.556
(std. dev.)	(0.153)	(0.177)	(0.193)	(0.306)	(0.436)	(0.583)	(0.494)	(2.049)
α	-0.931***	-0.796***	-1.152***	-0.988**	-0.575***	-0.550***	-0.459***	-0.625*
	(0.145)	(0.220)	(0.428)	(0.430)	(0.123)	(0.151)	(0.108)	(0.365)
$\overline{FDI}_{i,k}$	2.622***	1.305***	2.524***	5.138***	-0.744***	-0.675**	-0.587**	-5.587***
	(0.240)	(0.301)	(0.349)	(0.750)	(0.136)	(0.273)	(0.278)	(1.058)
Sector FE	Included	Included	Included	Included				
N	54	54	54	54	54	54	54	54
L	-22.31	-23.73	-21.55	-23.29	-2.76	-35.28	-33.46	-103.71
ρ					0.76	0.51	0.26	0.18
LR Test 1					50.48***	20.40**	5.95**	3.18***
LR Test 2	0.18	1.53	0.49	0.00				

The table shows the pooled regression of Equations (4.14a) and (4.14b) on different subsamples. Reactions of sectors to large terrorist attacks in regions are regressed on the share of FDI they receive. $\bar{\eta}_{i,k}$ is estimated using fractional logit, whereas $\bar{\zeta}_{i,k}$ is estimated using a random effects Tobit model. Standard errors are in parentheses, and are robust for $\bar{\eta}_{i,k}$. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent. L is the log likelihood value, ρ is the fraction of variance due to the unobserved heterogeneity μ_i and is defined as $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\epsilon^2}$. LR Test 1 is a likelihood ratio test of $\sigma_\mu = 0$ in $\bar{\zeta}_{i,k}$. LR Test 2 is a likelihood ratio test for the joint significance of the sector fixed effects in $\bar{\eta}_{i,k}$.

Table 4.A.4: Regression output - Variable TER

	A. Jump probability $\eta_{i,k,t}$					B. Jump size $\zeta_{i,k,t}$				
	$TER = 1$	$TER = 3$	$TER = 5$	$TER = 7$	$TER = 9$	$TER = 1$	$TER = 3$	$TER = 5$	$TER = 7$	$TER = 9$
mean	0.289	0.295	0.317	0.353	0.391	-0.560	-0.587	-0.718	-1.002	-1.460
(std. dev.)	(0.155)	(0.174)	(0.240)	(0.316)	(0.374)	(0.544)	(0.614)	(1.086)	(2.054)	(3.455)
α	-0.817***	-0.553***	-0.730***	-0.917***	-0.956***	-0.526***	-0.605***	-0.814***	-1.235***	-1.922***
	(0.05)	(0.126)	(0.183)	(0.229)	(0.263)	(0.101)	(0.112)	(0.175)	(0.307)	(0.503)
FDI$_{i,k,t}$	-0.115	0.389***	1.203***	1.723***	1.849***	0.046	-0.223***	-1.103***	-2.539***	-4.419***
	(0.076)	(0.124)	(0.202)	(0.270)	(0.315)	(0.057)	(0.070)	(0.161)	(0.331)	(0.573)
Year FE	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Sector FE	Included	Included	Included	Included	Included					
N	696	696	696	696	696	696	696	696	696	693
L	-264.95	-270.12	-291.66	-325.27	-361.66	-161.85	-312.97	-883.03	-1380.07	-1749.09
ρ						0.46	0.39	0.17	0.09	0.07
LR Test 1						391.23***	310.45***	105.61***	48.88***	32.74***
LR Test 2	7.11	6.04	3.04	1.59	1.50					

The table shows the panel regression of Equations (4.14a) and (4.14b) for different levels of TER. $\eta_{i,k,t}$ is estimated using fractional logit, whereas $\zeta_{i,k,t}$ is estimated using a random effects Tobit model. Standard errors are reported in parentheses, and are robust for $\eta_{i,k,t}$. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent. L is the log likelihood value, ρ is the fraction of variance due to the unobserved heterogeneity μ_i and is defined as $\frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma_\epsilon^2}$. LR Test 1 is a likelihood ratio test of $\sigma_\mu = 0$ in $\zeta_{i,k,t}$. LR Test 2 is a likelihood ratio test for the joint significance of the sector fixed effects in $\eta_{i,k,t}$.

Appendix 4.B Industry classifications and Foreign Direct Investment

The data on the investment position of U.S. sectors were obtained from the website of the U.S. Bureau of Economic Analysis in the section 'Position on a historical-cost basis, country detail by selected industry' and include all countries in which there is direct investment.

Industries are classified using the Standard Industrial Classification (SIC) prior to 1999 and according to the North American Industry Classification System (NAICS) thereafter. The industries for which data are publicly available and their respective SIC/NAICS codes are shown in Table 4.B.1.

Table 4.B.1: Industry classifications

Industry name	SIC	Industry name	NAICS
Oil and gas extraction		Mining	21
+ Petroleum and coal products	13 + 29	Utilities	22
Manufacturing, of which		Manufacturing, of which:	
Food and kindred products	20	Food	311
Chemicals and allied products	28	Chemical	325
Primary and fabricated metal industries	33-34	Primary and fabricated metal products	331-332
Industrial machinery and equipment	35	Machinery	333
Electronic and other electric equipment	36	Computers and electronic products	334
Transportation equipment	37	Electrical equipment, appliances, and components	335
Miscellaneous manufacturing industries	39	Transportation equipment	336
		Miscellaneous	339
Wholesale trade	50-51	Wholesale trade	42
		Information	51
Depository institutions	60	Depository institutions	60
Financial, insurance, and real estate industries	61-67	Finance and insurance	52
Services	70-89	Professional, scientific, and technical services	54
		Holding companies (nonbank)	55
Other	n.a.	Other	n.a.

The table displays the two industry classification available from the U.S. Bureau of Economic Analysis data on outward FDI of U.S. multinational firms. Prior to 1999, the data are classified using SIC, afterwards they are recorded using NAICS.

The BEA also provides a category Other Industries, which combines all remaining industries. Since this category can not be mapped into a series of SIC or NAICS codes we are forced to exclude it. Another exclusion is the Utilities category, as it is only available up until 2002. For regions where Miscellaneous Manufacturing is missing, we use the provided category Total Manufacturing and subtract all the available separate manufacturing categories.

To match the SIC/NAICS data on investments to the GICS-based stock market indices, we download all current and historic S&P500 companies with their SIC/NAICS and GICS codes from Compustat. We tabulate the SIC/NAICS classifications per 2 digit GICS code and obtain the mapping as shown in Table 4.B.2.

We observe from the mapping that the SIC/NAICS sectors do not correspond one-to-one with their GICS counterparts. For example, firms that are classified as Wholesale Trade appears in Consumer Staples, Health Care and Materials. Another NAICS sector that appears in multiple GICS sectors is Information, mapped into GICS counterparts IT, Telecom and Consumer Discretionary.

Under the SIC classification, this sector was unavailable and therefore we were not able to obtain an estimate of outward U.S. FDI for the Telecom sector in 1998.

Outside of Telecom in 1998 and the non-availability of the Utilities sector, other GICS sectors do not suffer from this problem. Unfortunately however, we do not have more disaggregate data at our disposal to map the SIC/NAICS sectors more accurately to their GICS counterpart. The 'Holding Companies (nonbank)' and 'Other' categories do not lead to a clear SIC/NAICS mapping and therefore have to be excluded, although they account for 37 percent of yearly U.S. outward FDI on average.

The BEA data is available on region and country level, with the limitations that some country/industry/year combinations are not shown to avoid disclosing data of a specific firm. Since combinations of industry/region/year do not suffer from this limitation, we use the BEA regions Latin America, Middle East, Africa and Asia & the Pacific. The European Union is also reported and takes into account changes in the number of member states. The Rest of Europe is then defined as the value of Europe minus the European Union. Countries in the Rest of Europe are for example Albania, Armenia, Bulgaria, Croatia, Kazakhstan, Moldova, Norway, Russia, Serbia and Switzerland. Since Norway and Switzerland fit in better with E.U. countries, and outward FDI data is always available for both countries, we place them together with the E.U. countries.

Table 4.B.2: Mapping SIC/NAICS to GICS

GICS	1998 (SIC)	1999-2010 (NAICS)
10. Energy	Oil and gas extraction + Petroleum and coal products (13 + 29) Industrial machinery and equipment (35)	Mining (21), Machinery (333)
15. Materials	Chemicals and allied products (28), Primary and fabricated metal industries (33-34), Wholesale trade (50-51)	Primary and fabricated metal products (331-332), Chemical (325), Wholesale trade (42)
20. Industrials	Primary and fabricated metal industries (33-34), Industrial machinery and equipment (35), Electronic and other electric equipment (36), Transportation equipment (37)	Primary and fabricated metal products (331-332), Machinery (333), Computers and electronic products (334), Electrical equipment, appliances and components (335), Transportation equipment (336)
25. C. Discretionary	Electronic and other electric equipment (36), Transportation equipment (37), Services (70-89)	Transportation equipment (336), Miscellaneous (339), Information (51), Professional, scientific, and technical services (54)
30. C. Staples	Food and kindred products (20), Chemicals and allied products (28), Wholesale Trade (50-51)	Food (311), Chemical (325), Wholesale Trade (42)
35. Health Care	Chemicals and allied products (28), Electronic and other electric equipment (36), Wholesale trade (50-51)	Chemical (325), Computers and electronic products (334), Miscellaneous (339), Wholesale trade (42)
40. Financials	Financial, insurance, and real estate industries (61-67), Depository institutions (60)	Finance and insurance (52), Depository institutions (60)
45. IT	Industrial machinery and equipment (35), Electronic and other electric equipment (36), Services (70-89)	Computers and electronic products (334), Information (51), Professional, scientific, and technical services (54)
50. Telecom		Information (51)

The table displays the mapping of SIC/NAICS sectors to GICS sectors, based on classifications of current and historic S&P500 companies obtained from Compustat.

5

Depositor Discipline and Bank Failures in Local Markets During the Financial Crisis

5.1 Introduction

Bank regulators have emphasized the role of market discipline in Pillar 3 of the Basel II and III accords. Contrary to monitoring by the regulator, market discipline relies on stakeholders such as depositors to monitor and, if necessary, prevent excessive risk-taking by banks. However, following the crisis the question is whether market discipline can still be used as a tool in bank supervision (see e.g. Acharya et al., 2014). The events that transpired during the crisis have given depositors mixed signals regarding the status and safety of banks, as well as the need to monitor their riskiness.

On the one hand, increases in the deposit insurance limit and government interventions such as bailouts have weakened incentives for depositors to monitor banks. In order to prevent bank runs, the U.S. government temporarily increased the level of deposit insurance from $100,000 to $250,000 in 2008, an increase made permanent in the Dodd-Frank act. However, even uninsured depositors have been compensated recently, with the FDIC assuming all deposits for most bank failures since the IndyMac Bank failure.[1] Moreover, regulators have shown that they are also willing to intervene in the shadow banking market with the bailout of money market funds. Given these interventions, depositors have little incentives ex-ante to engage in active monitoring of their depository institutions, as they are likely to be bailed out no matter how risky their banks are.

On the other hand, there have been over 400 commercial bank failures in the United States since the beginning of the crisis. Previous studies have documented that experiencing events such as bank failures can lead to a wake-up call among depositors (see e.g. Martinez Peria and Schmukler, 2001; Karas et al., 2010, 2013; Iyer and Puri, 2012). This finding is motivated by increased risk aversion of depositors, and is consistent with an aggregate experience hypothesis documented by, e.g., Malmendier and Nagel (2011). A wake-up call entails renewed discipline being exerted after bank failures, as these events can make depositors of other banks aware that their deposits are potentially also at risk. While some recent evidence supports the notion that government interventions have weakened overall market discipline (Cubillas et al., 2012; Berger and Turk-Ariss, 2014), it remains unclear whether depositors have woken up to the risks posed by their banks. This chapter provides evidence that, despite government interventions, depositors exerted discipline on their banks during the crisis and that the aforementioned failures did wake them up.

Martinez Peria and Schmukler (2001) were the first to document a wake-up of depositors, by showing that they were more responsive to banks riskiness after periods of crises in Argentina, Chile, and Mexico. Interestingly, despite expansion of deposit insurance coverage, both insured and uninsured depositors increased their monitoring. Similarly, using data on Russian banks, Karas et al. (2010, 2013) find evidence of a wake-up call for insured and uninsured depositors. Moreover, while the introduction of deposit insurance weakens overall discipline, it does not completely eliminate the wake-up call. Finally, using deposit account-level data of a bank in India, Iyer and Puri (2012) document a bank run after the failure of an unrelated bank.

Most of the work documenting the effect of the recent crisis on market discipline presents evidence of discipline on a subset of banks. For instance, Berger and Turk-

[1] Since the IndyMac Bank failed in 2008, the FDIC's Failed Bank List and accompanying press releases state that for most failures, the FDIC and eventual acquirer assumed all deposits.

Ariss (2014) find that while overall discipline decreased in both the United States and Europe, this can be chiefly attributed to decreased discipline for large and listed banks. Their finding is consistent with moral hazard by depositors following government interventions. For an international sample of banks, Bertay et al. (2013) find increased discipline on systemically large banks consistent with a wake-up call. However, similar to Berger and Turk-Ariss (2014) they document an absence of market discipline on the systemically largest banks in the United States. It therefore seems that depositors distinguish between risks posed by banks, and that they intentionally choose to discipline some banks while ignoring the risks of others. Correa et al. (2012) support this notion by documenting a bank run on U.S. branches of European banks during the European sovereign debt crisis. Evidence on developing countries is provided by e.g., Oliveira et al. (2014) who find an absence of discipline on Too-Big-To-Fail banks in Brazil. Moreover, using banks in CEE countries, Hasan et al. (2013) do not find evidence of an overall wake-up call, although there is increased discipline on affiliates of Western-European banks.

This chapter investigates if the wake-up call has materialized for depositors in the United States, and whether they discipline banks during the crisis despite a weakening of incentives. In doing so, it makes two contributions to the existing literature. First, depositor behavior is analyzed in local banking markets. Previous studies on market discipline have used the bank entity as their fundamental unit of analysis, comparing banks on a national or even supra-national level. However, they overlook that the relevant market for most banks and depositors alike is the local banking market. While, increasingly, U.S. banks have nationwide activities, these are only few in absolute numbers. Despite the trend of deregulation and consolidation in the U.S., banks and banking markets tend to be fragmented for well-known historical reasons. Depositors, therefore, are generally more likely to have deposits at those banks that are active within their local market. Even though online-banking has made it possible to deposit at out-of-market banks, it is often seen as a complement, not a substitute, to physical bank branches (see e.g. DeYoung and Hunter, 2002; DeYoung, 2005; DeYoung et al., 2007; Hernando and Nieto, 2007; Onay and Ozsoz, 2013). Discipline by depositors can therefore be expected mainly on banks active in their local market, and only relative to the other banks that are also present there. For this reason, depositor behavior is best analyzed in local banking markets.

Second, by analyzing these local markets, this chapter presents a cleaner identification of a possible wake-up call during the financial crisis compared to previous studies. While those studies rely on changes in deposits between banks before and after the crisis, this chapter exploits the fact that banks have branches in multiple markets. Even though bank riskiness is determined and evaluated at the level of the bank, depositors can react differently to these risks across local markets. The same bank with branches operating in multiple markets can thus be subject to a varying degree of discipline in each market, depending on whether a failure occurs in the market. The identification strategy employed in this chapter uses the branches of the bank in markets without failure as a control, whereas the branches of the bank that operate in a market *with* a failure are used as the treatment group. Hence, the wake-up call is measured as the difference in depositor reaction across the markets in which the bank is active, where some have experienced a failure while others have not. If experiencing bank failures is indeed the channel through which a wake-up call is achieved, I expect to find an increase in discipline for those markets where the failing banks operated a branch compared to markets without a failure.

To perform this analysis, I use publicly available information on the level of deposits in bank branches in the U.S., obtained from the FDIC's Summary of Deposits. I show that depositor discipline was present at the level of the local market between 2007 and 2013. Moreover, a wake-up call indeed materializes in those markets witnessing a bank failure, and this effect does not die out after 1 year but is long-lasting. Finally, I find that depositors react differently to bank failures of banks that are considered mainly local, compared with failures of banks whose headquarters are located out-of-market or even out-of-state. As such, this chapter offers implications for the regulatory and supervisory set-up following the crisis. For instance, the Net Stable Funding Ratio (NFSR) introduced in Basel III will force banks to hold a sufficient amount of stable funding, which includes demand and other customer deposits. This chapter shows that deposit funding of banks depends not only on the bank's risk characteristics, but also on what is happening to other banks in their market. Moreover, this chapter also suggests that prompt action by the FDIC to close down failing banks intended to stop bank runs can actually lead to more involved depositors and possibly a safer banking system.

The rest of the chapter is organized as follows. Section 5.2 discusses the methodology and data, followed by Section 5.3 which presents the results. Robustness tests are performed in Section 5.4, after which I briefly conclude.

5.2 Data and methodology

5.2.1 Identification

Previous studies on market discipline of participants other than depositors have investigated whether publicly available risk indicators are priced into bank funding, signaling the perceived risk to banks. For example, they find that riskier banks have higher bond yields/spreads (see e.g. Avery et al., 1988; Flannery and Sorescu, 1996; Jagtiani et al., 1999; Jagtiani and Lemieux, 2001; Sironi, 2003; Ashcraft, 2008) and a lower market value of equity (see e.g. Billett et al., 1998; Park and Peristiani, 2007; Baele et al., 2014). Depositors, on the other hand, can discipline their banks in more ways than one (Flannery, 1994). Similar to (subordinated) debt holders, depositors can require a higher interest rate from riskier banks, thereby compensating them for the risk of losing their deposits in the event the bank fails. Moreover, since deposits are liquid, depositors can also withdraw them if they feel the bank is taking excessive risks.

To measure market discipline the following equation is usually estimated:

$$\Delta \ln D_{i,t} = \beta_0 + \beta_1 \mathbf{Risk}_{i,t-1} + \beta_2 \mathbf{Controls}_{i,t-1} + \beta_i + \beta_t + \epsilon_{i,t} \tag{5.1}$$

where $\Delta \ln D_{i,t}$ is the growth rate of deposits of bank i in year t, $\mathbf{Risk}_{i,t-1}$ is a vector of lagged indicators of bank risk, $\mathbf{Controls}_{i,t-1}$ is a vector of control variables, and β_i and β_t are bank and time fixed effects to control for unobserved heterogeneity. Market discipline is present if depositors move their deposits to safer banks. Therefore, safer banks are expected to have a higher growth rate in deposits compared to riskier banks. In Equation (5.1), a negative β_1 coefficient signals the presence of market discipline, as safer banks see higher deposit growth compared to more risky peers. Risk and control variables are lagged to avoid possible endogeneity, as, for instance, bank risk can be endogenous to depositor behavior. In the absence of a counterfactual, the wake-up call is identified as the difference in behavior before or after crises, attributing all changes in behavior to the crisis.

In this chapter, I propose to identify the wake-up call in a different manner. To measure the wake-up call, a perfect counterfactual would entail the same bank being exposed to a scenario with a crisis and one without. The difference in depositor behavior between scenarios would then directly identify the effect of a crisis, controlling for unobserved characteristics which impact both the control and the treated group. In the setting of depositor discipline, one would thus prefer to have the same bank in two different markets, but with different realizations of depositor behavior due to certain events. Previous literature suggests that the events leading to a wake-up call could be bank failures, as these can make depositors aware that their deposits are possibly at risk (see e.g. Martinez Peria and Schmukler, 2001; Karas et al., 2010, 2013).

While depositors can react differently across markets to this multi-market bank, bank risk itself is determined and evaluated at the bank-level. The FDIC, for instance, does not close down single branches but intervenes only in the bank entity. Identification of the wake-up call then relies on multi-market banks, which share the same bank risk across markets but experience different levels of discipline in markets with failures. This is visualized in Figure 5.1, where a bank is active in 7 markets, 3 experienced bank failures.[2] In light of the figure, the question is whether depositors were more aware of the risks of the bank in the markets with failure.

Since bank risk is determined at the level of the bank, the differences in depositor reaction between markets can only be due to the failures, using the markets without failures as a control group. Empirically, this effect is obtained by performing the following regression:

$$\begin{aligned}\Delta \ln D_{i,m,t} = {} & \beta_0 + \beta_1 \mathbf{Risk}_{i,t-1} + \beta_2 \mathbf{Controls}_{i,m,t-1} + \beta_3 F_{m,t-k} \\ & + \beta_4 F_{m,t-k} \mathbf{Risk}_{i,t-1} + \beta_5 F_{m,t-k} \mathbf{Controls}_{i,m,t-1} + \beta_i + \beta_m + \beta_t + \epsilon_{i,m,t} \quad (5.2)\end{aligned}$$

where $\Delta \ln D_{i,m,t}$ is the growth rate of deposits of bank i in market m at time t, $\mathbf{Risk}_{i,t-1}$

[2]For the purpose of this chapter, a market is considered to experience a failure through the presence of branches of banks that where closed down by the FDIC.

Figure 5.1: Identification strategy

is still a vector of lagged indicators of bank risk measured at the level of the bank, $\mathbf{Controls}_{i,m,t-1}$ is a vector of control variables that can differ between banks or markets, and β_i, β_m and β_t are bank, market and time fixed effects to control for unobserved heterogeneity. Moreover, $F_{m,t-k}$ is a dummy variable indicating whether a bank failure occurred in market m during the last k years and is interacted with the risk and control variables. In this setting, a negative and significant β_4 now measures whether the presence of a failed bank leads to excess market discipline, with the combined discipline effect being $\beta_1 + \beta_4 F_{m,t-k}$. Moreover, the length of the window k can be adjusted to measure short-run ($k = 1$) or long-run ($k > 1$) effects. If bank failures do indeed lead to a wake-up call, it follows that the change in behavior of depositors should be noticeable in markets that experience them compared to those that do not. In Equation (5.2), this effect is picked up by the β_4 coefficient.

Control variables are included to control for market-specific characteristics, such as the concentration and the number of branches present, or for bank-specific characteristics. To control for the fact that depositors might move their deposits to a risky bank as long as they are sufficiently rewarded with a higher interest rate, I include the bank-specific (implicit) lagged deposit interest rate. Previous studies have indeed found evidence that risk is priced priced into bank deposit rates (see e.g. Hannan and

Hanweck, 1988; Ellis and Flannery, 1992; Brewer III and Mondschean, 1994; Cook and Spellman, 1994), or that bank risk is priced into both the price and the quantity of deposits (see e.g. Park, 1995; Park and Peristiani, 1998; Demirgüç-Kunt and Huizinga, 2004; Ioannidou and de Dreu, 2006; Karas et al., 2010, 2013; Bertay et al., 2013). Put differently, riskier banks need to offer a higher interest rate to maintain or attract new deposit funding. In most cases however, the interest rate employed is an implicit interest rate, calculated as the total interest expense of a bank divided by total loans. The downside of this approach is that it does not separate between interest rates paid on insured or uninsured accounts, or between rates on existing accounts compared to new ones. Moreover, and more importantly in the context of this chapter, the implicit interest rate is bank-specific, making it difficult to investigate the presence of price discipline in *local* markets. If the deposit rate is determined at the bank level, price discipline in local markets would be incorrectly identified. This depends on whether multi-market banks set their interest rates uniformly, or whether they differ across markets, something on which there is no consensus as of yet. While Craig and Dinger (2013) find that there is cross-market variation in deposit rates of multi-market banks, this is not found by Radecki (1998), Heitfield (1999) and Park and Pennacchi (2009).[3] Since identification is potentially incorrect, I do not explicitly estimate a price equation, but instead follow the estimation strategy employed by, e.g., Maechler and McDill (2006), and use the lagged implicit interest rate to take into account that these rates can influence the quantity of deposits.

Equation (5.2) is estimated using OLS. The high-dimensionality due to the extra dimension m and the number of banks and markets involved means that many different types of fixed effects can be included. Besides the standard bank, market and time fixed effects for instance, it is possible to control for other sources of unobserved heterogeneity. For instance, besides market fixed effects to control for market-specific changes in deposits, and year fixed effects to control for countrywide business cycle effects, market$\times$year fixed effects pick up local market business cycle effects that could be misinterpreted as depositor discipline. Similarly, bank$\times$market fixed effects can be employed to control for market-specific strategies, as banks might want to expand

[3] Unfortunately, the deposit rate that banks choose to offer to depositors in a given regional market at a given time is not publicly available. A possible solution is to obtain the local price level from banks that are only active in that market. However, the identification strategy employed in this chapter depends on multi-market banks. A price level based on local banks would undermine this strategy.

their market share in certain markets but not in others. To be able to include these types of fixed effects, I implement a within estimator specifically designed for panel data models with high dimensionality. This estimator is preferred over the LSDV estimator, since the LSDV estimator can be computationally difficult to estimate due to the large number of dummy variables required. However, the unbalanced nature of the dataset has to be taken into account. Mátyás and Balázsi (2012) show that applying an analytical within-transformation developed for a balanced panel leads to a bias in an unbalanced setting that does not drop out as $N \rightarrow \infty$. Moreover, dynamic panel models also suffer from this bias when dealing with an unbalanced setting. To estimate unbalanced panel models with high-dimensional fixed effects, Guimarães and Portugal (2010) therefore propose a feasible iterative approach based on demeaning the (in)dependent variables over one dimension at a time, in order to minimize the regression RMSE. While Guimarães and Portugal (2010) implement this iterative procedure for 2 fixed effect dimensions, it can easily be extended to 3 (Torres et al., 2013) or even N fixed effects (Rios-Avila, 2013).

To obtain standard errors for the original model, Rios-Avila (2013) suggests to use the algorithm proposed by Abowd et al. (2002), which is necessary to correct for the degrees of freedom.[4] The standard errors are clustered using the banking market as cluster. As OLS standard errors could be too small for proper inference testing in the case of within-cluster serial correlation, Bertrand et al. (2004) and Colin Cameron and Miller (2014) suggest using cluster-robust standard errors to account for possible serial correlation. Following the suggestion of Colin Cameron and Miller (2014), standard errors are clustered over the local markets as these are the fewest in number and lead to the highest standard errors.[5]

Next, I will describe the data used in this chapter, before proceeding with the presentation of the results and robustness checks.

5.2.2 Data

The data used in this analysis are obtained from multiple sources. The main source is the FDIC's Summary of Deposits (SOD), a publicly available annual survey re-

[4] I thank Fernando Rios-Avila for sharing his code.

[5] Robustness tests with clustering at the level of the bank were also performed and lead to smaller standard errors. Results are available upon request.

porting branch-level deposits for all FDIC-insured institutions, including insured U.S. branches of foreign banks. The FDIC requires all institutions with a main office and one or more branches to file this survey. While banks can choose how to assign deposits to their offices (e.g. by proximity to the address of the account holder, the main activity, or origination of the deposit account), it should be consistent with existing internal record-keeping practices (FDIC, 2014). All branches and banks in the SOD have a unique identifier and are geo-coded, meaning that the deposits of a bank can easily be allocated to each banking market where a branch office is located. Consistent with previous literature, local markets are defined as Metropolitan Statistical Areas (MSAs) and non-MSA counties (see e.g. Prager and Hannan, 1998; Berger et al., 1999; Collender and Shaffer, 2003; Adams et al., 2007), although robustness tests are performed by estimating Equation (5.2) on each separate group and using the county as a local market. If multiple branches of the same bank are present in a market, the deposits located at these branches are consolidated to bank-market deposits. Since, as we will see later, the bulk of the bank failures occurred from 2007 onwards, I obtain data for the years 2007 - 2013.

Data used to calculate bank risk indicators are obtained from the Call Reports for Income and Condition, and linked to the SOD data using the bank's unique FDIC assigned certificate number. Unit banks, or other banks that have not filed the SOD survey, are subsequently dropped. Data on bank failures are obtained from the FDIC's Failed Bank List, which lists failed banks for which the FDIC is appointed as receiver, the closing date and, if available, the acquiring entity. To construct a variable indicating whether a failed bank operated a branch-office in a certain market, I combine the Failed Bank List with the SOD in the year prior to the failure, as banks do not file the SOD during the failure-year itself. To remove spurious increases in bank-market deposits, I remove bank-market observations when the bank has merged or acquired another bank that was active within the same market. Data on mergers and acquisitions are obtained from the Failed Bank List and the bank merger database maintained by the Federal Reserve Bank of Chicago.[6]

Since the SOD is an annual survey, collected on June 30$^{\text{th}}$, I use the second quarter (June) Call Report to match bank-market deposits to bank fundamentals. Years are

[6] Available at `http://www.chicagofed.org/webpages/publications/financial_institution_reports/merger_data.cfm`.

subsequently defined as periods between the filing moments and failure years are appropriately updated. For instance, a bank that failed on June 19th, 2008 is linked to the 2007 SOD. However, a bank that failed on July 20th, 2008 is linked to the 2008 SOD in order to determine where it was operating branches.

The SOD only reports total deposits and does not distinguish between insured and uninsured deposits. Ex-ante, most discipline can be expected from uninsured deposits as these better incentives to monitor the safety of their banks compared to insured depositors. However, previous research has shown that insured depositors are also capable of discipling banks, albeit to a lesser extent than uninsured depositors (see e.g. Park and Peristiani, 1998; Martinez Peria and Schmukler, 2001; Karas et al., 2013). Moreover, government interventions have blurred the lines between insured and uninsured deposits. First, the Dodd-Frank Act permanently increased the level of deposit insurance coverage from \$100,000 to \$250,000, thereby effectively insuring previously uninsured deposits.[7] Second, after the failure of IndyMac bank in 2008, the FDIC has assumed all deposits for most bank failures, effectively insuring the uninsured depositors. Third, the government has also bailed out the money market mutual funds, which are part of the shadow banking system and did not fall under deposit insurance in the first place. Given these interventions, it is safe to assume that uninsured depositors can behave as if they have insurance, since they are practically assured they will be compensated in the case of a failure. As all depositors were technically insured, the analyses provides a lower bound of the level of market discipline present in the local markets. Although this implicit deposit insurance could lead to less market discipline, studies on the recent financial crisis have shown that market discipline was not eliminated because of these bailout guarantees (see e.g. Gropp et al., 2011).

The final dataset consists of 6,735 banks that are active in 2,328 distinct markets (388 MSAs and 1,940 non-MSA counties) during a maximum of 7 years. The average number of markets served by banks in the sample is 3, although banks such as Bank of America, Wells Fargo and U.S. Bank have branches in more than 400 markets. Moreover, on average there are 6 banks present in a banking market, although some large MSAs like Minneapolis-St. Paul-Bloomington, Chicago-Naperville-Joliet,

[7] While the Act was only signed into federal law in July, 2010, Congress had approved a temporary increase in the deposit insurance limit starting on October 3, 2008. The Dodd-Frank Act retroactively increased the limit to also cover failures between January 1, 2008 and October 3, 2008.

or Dallas-Fort Worth-Arlington are served by more than 100 banks.

Summary statistics for the dependent and explanatory variables are reported in Table 5.1. While most of the variables of interest are ratio's, variables that are in levels are deflated to 2007Q2 dollars. The dependent variable, $\Delta \ln D_{imt}$, is the percentage change in deposits of bank i in market m in year t. The average bank-market experienced an increase in deposits of 1.73% per year during the sample period. Despite the disaggregation of the data across markets, the deposit growth mimics the values in Berger and Turk-Ariss (2014), who use bank-level data. The most a single bank-market loses in deposits from year-to-year is 29%, while the most deposits attracted represented a 58% gain. The main explanatory variables for bank risk are Equity (Equity/Total Assets), NPL (Non-Performing Loans/Total Loans), and ROA (Net Profit/Total Assets). These variables cover different dimensions of risk, as the capital buffer indicates the absorptive capacity of a bank to incur future losses, non-performing loans are a good indicator how large those future losses can be, and ROA indicates how fast a bank can rebuild the capital base when it retains earnings. F is a dummy variable indicating whether a failed bank was operating a branch in the local market. In the analysis, this variable is used to determine whether a failure has taken place in the market during the last k years. Control variables include business model characteristics (Real Estate Loans/Total Loans), efficiency estimates (Cost/Income), the ratio of liquid assets to total assets, securities held for sale and investment to total assets, the implicit interest rate calculated as the interest expense divided by total deposits, a dummy variable indicating whether the bank is a savings banks or a member of a Bank Holding Company, and market level characteristics such as the HHI (based on the level of deposits) and the number of branches present in the market. To remove outliers, the variables $\Delta \ln D_{imt}$, Equity, NPL, ROA, Real Estate/Total Loans, Cost/Income, Liquid Assets and Implicit Interest Rate are truncated at the 2.5 and 97.5 percentile. The summary statistics presented here are based on bank-market observations, banks operating in more markets are thus given a higher weight for those variables calculated at the level of the bank.

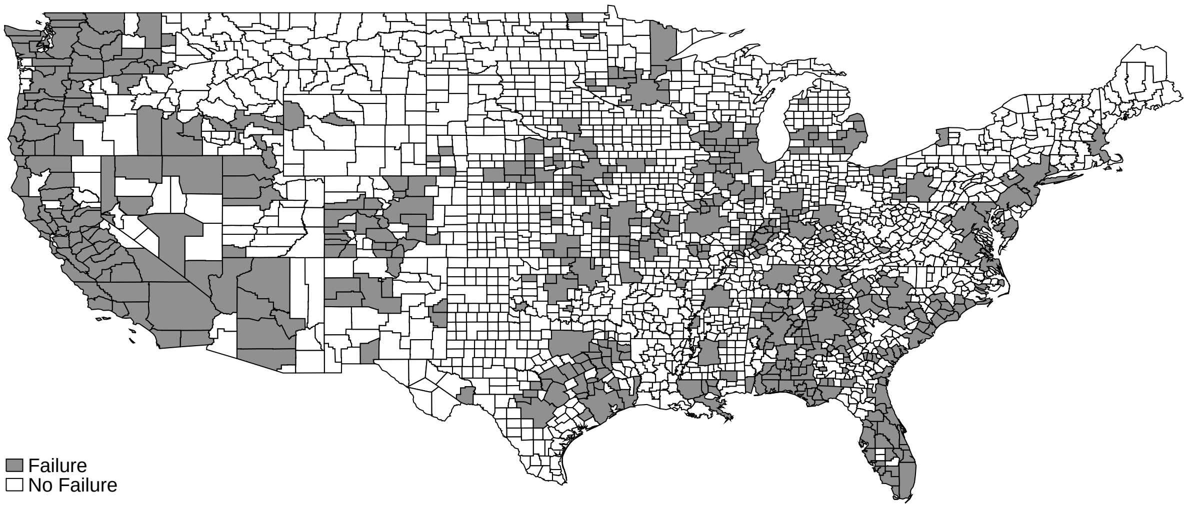
Figure 5.2 shows the geographical distribution of failures occurring during the sample period. From this figure, we see that most failures occur in and around large metropolitan areas, although they are divided between both metropolitan areas and

Table 5.1: Summary statistics

Variable	Obs	Mean	Std. Dev	Min	Max
Dependent Variable					
$\Delta \ln D_{imt}$	123615	0.0173	0.1370	-0.2949	0.5845
Main Explanatory Variables					
Equity	133117	0.1072	0.0251	0.0631	0.1960
NPL	136667	0.0149	0.0139	0	0.0698
ROA	133238	0.0038	0.0038	-0.0137	0.0122
F	140170	0.1513	0.3583	0	1
Control Variables					
Real Estate/Total Loans	132765	0.6917	0.1492	0.3131	0.9604
Cost/Income	133108	0.6892	0.1428	0.4355	1.2804
Liquid Assets (%)	133193	0.0802	0.0608	0.0148	0.3127
Securities for Sale and Investment	133172	0.1943	0.1060	0.0088	0.5185
Implicit Interest Rate	133156	0.0085	0.0057	0.0012	0.0228
Ln(Total Assets)	140170	14.079	3.0107	4.3474	21.201
Savings Banks	140170	0.0466	0.2108	0	1
BHC	140170	0.8799	0.3250	0	1
HHI	140170	0.2171	0.1398	0.0495	1
Ln(no. branches)	140170	0.8066	0.9322	0	6.9392

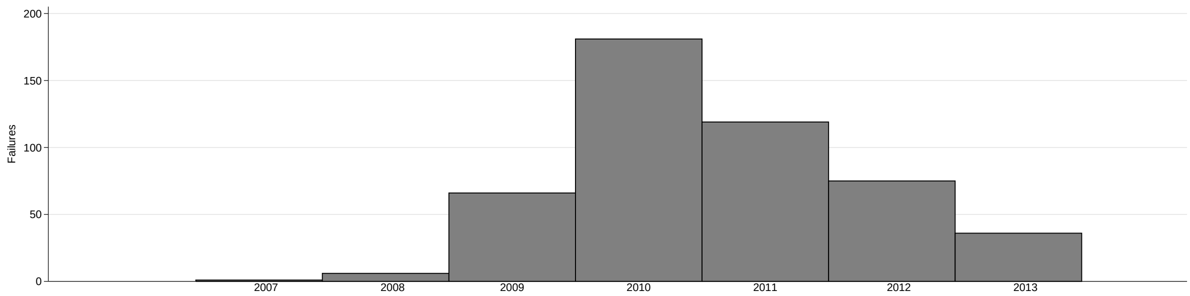
This table shows summary statistics for the variables used in the analysis. The dependent variable $\Delta \ln D_{imt}$ is the percentage change in deposits of bank i in market m in year t. The main explanatory variables are Equity (Equity/Total Assets), NPL (Non-Performing Loans/Total Loans) and ROA (Net Profit/Total Assets). F is a dummy variable indicating whether a local market has experienced a failing bank. The control variables include business model characteristics (Real Estate Loans/Total Loans), efficiency estimates (Cost/Income), the ratio of liquid assets to total assets, securities held for sale and investment to total assets, the implicit interest rate calculated as the interest expense divided by total deposits, a dummy variable indicating whether the bank is a savings banks or a member of a Bank Holding Company, and market level characteristics such as the HHI (based on the level of deposits) and the number of branches present in the market. For banks merging or acquiring other banks, bank-market observations of the dependent variable are removed. Furthermore, the variables $\Delta \ln D_{imt}$, Equity, NPL, ROA, Real Estate/Total Loans, Cost/Income, Liquid Assets and Implicit Interest Rate are truncated at the 2.5 and 97.5 percentile to exclude outliers. The data is deflated to 2007Q2 dollars.

Figure 5.2: Location of bank failures 2007 - 2013



This figure shows the location of all bank failures between 2007 and 2013 on a market-level. Markets, defined as a MSA or non-MSA county, where failed banks operated branches are highlighted. Source: FDIC Failed Bank List and Summary of Deposits.

Figure 5.3: Bank failures over time



This figure shows the number of bank failures in the U.S. per year. Source: FDIC Failed Bank List.

rural counties. Furthermore, Figure 5.3 plots the timing of the bank failures. It shows a clear peak in 2010 with over 150 banks failing, before steadily decreasing in the following years.

Table 5.2 shows summary statistics of failed and acquiring banks, and reports results of a t-test comparing the average differences. In total, there are 426 failing banks and 403 acquiring banks that can be matched to the SOD and Call Reports. The number of failing banks is higher than the number of acquiring banks because not all banks that were received by the FDIC were later sold, and because some banks acquired multiple failing banks. On average, acquiring banks were better capitalized, had a lower

Table 5.2: Difference failed and acquiring banks

	Failed Banks	Acquiring Banks	Difference Failed - Acquiring
Equity	0.039	0.121	-0.082^{***}
Loans	0.719	0.669	0.050^{***}
NPL	0.102	0.021	0.081^{***}
Deposits	0.888	0.782	0.106^{***}
Ln(Total Assets)	12.36	14.20	-1.840^{***}
BHC	0.815	0.916	-0.101^{***}
Savings	0.028	0.022	0.006
Number of Markets	7.525	183.8	-176.262^{***}
N	426	403	

This table shows differences in some key statistics between the failed banks and the banks that acquired them. The variables are Equity (Equity/Total Assets), Loans (Loans/Total Assets) NPL (Non-Performing Loans/Total Loans), Deposits (Deposits/Total Assets), Ln(Total Assets), variables indicating whether the banks were part of a Bank Holding Company (BHC) or were a Savings bank (Savings), and the average number of markets it had branches in. T-tests to compare the average were performed and reported in the last column. The number of failed banks differs from the number of acquiring banks because not all failed banks were acquired, and some banks acquired more than 1 failed bank.

share of non-performing loans, less loans overall, and were less reliant on deposits for their funding. The failed banks were also significantly smaller, with their assets totaling $233 million versus $1.4 billion for the acquiring banks. Moreover, acquiring banks were more likely to be a member of a Bank Holding Company, while the number of savings banks that failed or acquired other banks was relatively low at around 2%. Finally, failed banks operated branches across on average 7 local markets, while the acquiring ones were active in 183 markets. These numbers, however, are skewed because of a few large failures and subsequent acquisition. Median values - 4 and 18 markets for failed and acquiring banks, respectively - also confirm that acquiring banks were active in more markets.

Finally, Table 5.3 shows the characteristics of local markets with and without a failure. In general, when considering both MSAs and non-MSA counties, we can see that markets with failures were characterized by a lower level of concentration and more branches, deposits and banks. The year after failure, banks active in non-failed markets experienced higher deposit growth. The same conclusions hold when considering only the largest markets (MSAs), albeit to a lesser extent. In general, as indicated by

Table 5.3: Difference markets with and without a failure

Panel A: MSA and non-MSA counties

	Market level				Bank-level		
	Non-Failed	Failed	Difference		Non-Failed	Failed	Difference
HHI	3484	2047	1437***	$\Delta \ln D_{i,m,t}$	0.020	0.002	0.018***
Ln(Number of Branches)	2.352	3.964	-1.612***				
Ln(Total Deposits)	12.69	14.66	-1.973***				
Ln(Number of Banks)	1.666	2.698	-1.033***				

Panel B: MSA

	Market level				Bank-level		
	Non-Failed	Failed	Difference		Non-Failed	Failed	Difference
HHI	1708	1526	182***	$\Delta \ln D_{i,m,t}$	0.033	0.006	0.028***
Ln(Number of Branches)	4.431	5.156	-0.725***				
Ln(Total Deposits)	15.14	16.10	-0.954***				
Ln(Number of Banks)	2.860	3.365	-0.505***				

This table shows differences in some key statistics between markets with and without failed banks. The market-level variables are the deposit-level Herfindahl-Hirschman Index, the number of branches operating in the market, the total deposits available in the market and the total number of banks operating in the market. The bank-level variable is the dependent variable, the change in the level of deposits of bank i in market m at time t. This bank-level variable is measured the year after failure, while the market-level characteristics are measured during the year of failure. Panel A shows the statistics for both MSA and non-MSA counties, while Panel B does this for only MSAs.

Figure 5.2, bank failures occur in larger markets. To ensure that the results are not driven by this, I perform robustness tests in Section 5.4 using only the larger markets.

5.3 Results

This section present the results of the analysis. First, I test for the existence of overall market discipline in local markets during the crisis by analyzing whether banks with better risk profiles see a higher growth in their deposits. In a second step, I test whether local market with a bank failure experience extra discipline from depositors. Finally, I explore some outcomes of the underlying mechanism to provide additional evidence, before concluding with several robustness tests.

5.3.1 Baseline specification

The results from the analysis of Equation (5.2) are presented in Table 5.4. I report the coefficients for the main risk characteristics, and refer the reader to Table 5.A.1 for the full results. In all columns, the dependent variable is the change in deposits of bank i in market m in year t, while the variables of interest are the lagged capital ratio (Equity), the share of Non-Performing Loans (NPL) in the total loan portfolio, and the Return-on-Assets (ROA). All specifications include bank, market and year fixed effects.

In column (1), we can see that market discipline was present in local markets throughout the sample, as the coefficients on the risk characteristics are statistically significant and have the correct sign. Branches belonging to banks that reported higher equity levels, a lower share of non-performing loans, and a higher return-on-assets in the previous year see a higher growth of deposits compared to other banks active within the market. When comparing banks at the 75th and 25th percentile across the three risk variables, I find that the economic effect of an improvement in the capital buffer is most important. A bank at the 75th percentile of the equity ratio distribution experiences an extra deposit growth of 2.5% compared to a bank at the 25th percentile. This value is -1.3% for the non-performing loans and 0.9% for an improvement in profitability. These are relatively large effects considering that the average deposit growth of banks in local markets is 1.7%.

In columns (2)-(4), the risk and control variables are interacted with $F_{m,t-k}$, a dummy variable indicating whether a bank failure occurred in market m during the last k years. Column (2) measures the excess discipline when a failure occurred in the preceding year, column (3) does the same for 2 years since failure, and column (4) measures the excess discipline if the time since failure is 3 or more years. Including the interactions does not change the sign or significance of the main explanatory variables, but it does show that excess discipline is exerted on banks in local market with failures. In the first year after bank failure (column (2)), we can see that extra discipline is mainly directed at the capital ratio of banks and at the share of non-performing loans. In the years afterwards, discipline is mainly directed at the capital ratio of banks, although the return-on-assets also becomes statistically significant. Regarding the economic significance of these coefficients, the capital ratio is again by far the most significant. In years 2 and 3+ after failure, the difference in deposit growth between a bank at the 75th and 25th percentile of the capital ratio is 3%, while this is only 1.2% for the return-on-assets.

Table 5.4: Do failures lead to a wake-up effect?

	(1)	(2)	(3)	(4)
	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$
Time since failure (k)		1 year	2 year	3+ year
Equity	0.853***	0.801***	0.790***	0.784***
	(0.063)	(0.058)	(0.058)	(0.059)
NPL	−0.725***	−0.657***	−0.670***	−0.683***
	(0.064)	(0.067)	(0.067)	(0.068)
ROA	2.435***	2.277***	2.091***	1.920***
	(0.264)	(0.286)	(0.293)	(0.298)
F		−0.086***	−0.066***	−0.076***
		(0.028)	(0.026)	(0.025)
$F\times$ Equity		0.266***	0.206***	0.196***
		(0.065)	(0.060)	(0.058)
$F\times$ NPL		−0.299**	−0.166	−0.123
		(0.120)	(0.111)	(0.108)
$F\times$ ROA		0.846	1.142**	1.433***
		(0.550)	(0.510)	(0.488)
Control Variables	Included	Included	Included	Included
Bank Fixed Effects	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included
Within $\bar{R}^2$	0.031	0.032	0.033	0.034
N	84,131	84,131	84,131	84,131

This table shows the estimation results for Equation (5.2), regressing the percentage change in deposits of bank i in local market m in year t on risk measures of bank i in year $t-1$. Control variables are as indicated in Table 5.1, every specification includes bank, market and time fixed effects to control for demand effects. Column (2) measures the impact on market discipline 1 year after the failure has occurred, column (3) measures the long-run effect 2 years after failure and column (4) measures the long-run effect 3 or more years after bank failure. Full results are shown in Table 5.A.1. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

The results presented here suggest that depositor discipline was present in local markets during the crisis, and that experiencing bank failures leads to a wake-up call of depositors. However, interpreting the results in this manner can be misleading. Between 2009 and 2011, the Deposit Insurance Fund essentially ran out of money as evidenced by its negative reserve ratio (FDIC, 2012). To help the FDIC maintain its credibility, Congress created an emergency borrowing line in March, 2009 of up to $100 billion, with a possibility to increase this emergency line to $500 billion conditional on the approval from the Federal Reserve Board and the Treasury. Combined with the fact that most failures occurred in 2010 and 2011 (see Figure 5.3), the result in Table 5.4 might actually be due to an aggregate change in depositor behavior and not to bank failures. To explore this possibility, I split up the sample for the years 2008 - 2010 and 2011 - 2013, and estimate Equation (5.2) on these subsamples. The results are shown in Table 5.5. Specification (1) estimates Equation (5.1) using the standard bank, market and year fixed effects, and shows that general discipline is present during both crisis-periods. Moreover, the coefficients are of the same magnitude, indicating no structural break in behavior as the crisis progresses. The magnitude of the coefficients on NPL and the implicit interest rate do change, but this can be attributed to the change in the average of the variables. Whereas the average equity ratio and the average return-on-assets remained stable around 10% and 0.37%, the NPL and the implicit interest rate changed from 1.1% to 1.9%, and from 0.5% to 1.4% in the latter stages of the crisis, respectively. Finally, note that since bank fixed effects are included, the identification on the BHC and Savings Bank variables relies on banks becoming BHC members and savings banks becoming commercial banks, or vice versa.

5.3.2 Do failures in all markets lead to extra discipline?

Having established that bank failures give an extra incentive to depositors to monitor banks active in their local market, I explore several outcomes that can be the consequence of this channel. The first is whether depositors receive different signals from different bank failures. For instance, a failing bank with a high local presence is presumably a stronger signal to depositors than a bank with little local presence. Since the former failure is likely due to a downturn in the local business cycle, depositors

Table 5.5: Is there a difference during the crisis?

	(1) $\Delta \ln D_{imt}$		(2) $\Delta \ln D_{imt}$		(3) $\Delta \ln D_{imt}$	
	2008 - 2010	2011 - 2013	2008 - 2010	2011 - 2013	2008 - 2010	2011 - 2013
Main Explanatory Variables						
Equity	0.880***	0.919***	0.961***	0.902***	0.979***	0.948***
	(0.102)	(0.126)	(0.101)	(0.126)	(0.114)	(0.140)
NPL	−0.765***	−0.481***	−0.708***	−0.486***	−1.031***	−0.475***
	(0.131)	(0.106)	(0.129)	(0.103)	(0.145)	(0.117)
ROA	1.513***	1.600***	0.872*	1.688***	2.231***	1.603***
	(0.475)	(0.434)	(0.477)	(0.436)	(0.532)	(0.492)
Control Variables						
Implicit Interest Rate	1.527***	4.560***	0.975*	5.226***	0.205	3.862***
	(0.537)	(1.028)	(0.531)	(1.028)	(0.623)	(1.157)
Ln(Total Assets)	−0.130***	−0.157***	−0.138***	−0.159***	−0.150***	−0.171***
	(0.014)	(0.012)	(0.014)	(0.013)	(0.015)	(0.014)
BHC	−0.027*	−0.012	−0.024*	−0.026*	−0.031**	−0.022
	(0.014)	(0.015)	(0.014)	(0.015)	(0.014)	(0.017)
Ln(Number of Branches)	−0.013***	−0.006***	−0.028***	−0.030***		
	(0.001)	(0.001)	(0.011)	(0.010)		
HHI	0.064	0.010	0.067	0.011		
	(0.040)	(0.029)	(0.043)	(0.029)		
Cost/Income	−0.006	0.037**	−0.015	0.035**	−0.014	0.033*
	(0.017)	(0.016)	(0.017)	(0.016)	(0.019)	(0.017)
Real Estate Loans	−0.104***	−0.048*	−0.092***	−0.049*	−0.056*	−0.051*
	(0.030)	(0.026)	(0.030)	(0.026)	(0.032)	(0.028)
Liquid Assets	−0.591***	−0.361***	−0.580***	−0.361***	−0.533***	−0.351***
	(0.034)	(0.029)	(0.033)	(0.028)	(0.037)	(0.031)
Securities for Sale and Investment	−0.264***	−0.288***	−0.247***	−0.286***	−0.260***	−0.292***
	(0.026)	(0.024)	(0.026)	(0.024)	(0.028)	(0.027)
Savings Bank	0.215**	−0.100***	0.207**	−0.101***	0.262**	−0.097***
	(0.100)	(0.005)	(0.098)	(0.005)	(0.107)	(0.005)
Bank Fixed Effects	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included
Bank×Market Fixed Effects			Included	Included	Included	Included
Market×Year Fixed Effects					Included	Included
Within $\bar{R}^2$	0.038	0.032	0.050	0.048	0.056	0.047
N	40,186	43,945	40,186	43,945	40,186	43,945
Failed	0.138	0.168	0.138	0.168	0.138	0.168

This table explores the difference in depositor reaction during two phases of the crisis using a subsample analysis. Specification (1) shows the results for the years 2008 - 2010 and the years 2011 - 2013. Specifications (2)-(3) repeat the analysis but add bank×market and market×year fixed effects, respectively. Since 'HHI' and 'Ln(Number of Branches)' are market-level indicators and vary only over market and time, they are dropped in specification (3). 'Failed' indicates the percentage of markets that experienced a failure during the time period. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

might become more risk-averse after a local failure (analogous to Malmendier and Nagel, 2011). If bank failures indeed lead to a wake-up call because of a downturn in macro-economic conditions, local failures should give a stronger signal and incentive to depositors to increase their monitoring. To investigate this possibility, I divide the failures into banks with headquarters in-market and banks whose headquarters are located out-of-market. The results are presented in Table 5.6, and show that for both in-market and out-of-market failures depositors monitor the equity ratio in the short and long run. Interestingly, and consistent with the experience hypothesis, depositors monitor the NPL of their banks for 3+ years after *in*-market failures but not for *out*-of-market failures. Moreover, while the return-on-assets is significant in later years after bank failures, this is only the case for out-of-market failures. These results indicate that local bank failures indeed give a stronger signal than non-local bank failures. Since local bank failures are presumably caused by local business cycle effects, I interpret this as extra evidence for the failure-induced wake-up call. Results for a similar exercise, where I look at failed banks with their headquarters in the same state, or headquarters in another state than the local market, are reported in Table 5.A.2. The results are quantitatively similar and are therefore not discussed here.

Another question is whether depositors wake up to the same extent when they have less choice to move their deposits within the same market. In markets with few banks, or highly concentrated markets, where is the depositor going to run to? If depositors have no choice between banks, they are less likely to move their deposits, in effect disciplining banks less. A bank failure is thus expected to lead to a lower depositor reaction in concentrated markets compared to less concentrated markets. I split up the markets according to their deposit-based HHI, following the U.S. Department of Justice guidelines to divide banking markets into concentrated markets (HHI ≥ 2500), moderately concentrated markets ($1500 \leq$ HHI < 2500) and markets with a low level of concentration (HHI < 1500). The results are shown in Table 5.7. General discipline is present in all markets, but depositors react differently across markets when faced with a bank failure. In markets with a high level of concentration, depositors focus mainly on the return-on-assets to decide where to move their deposits to. In less concentrated markets, depositors tend to value the equity ratio, which is arguably a better indicator of the risk of a bank than the return-on-assets. Depositors seem to

Table 5.6: Is there a difference between in-market and out-of-market failures?

	(1)		(2)		(3)	
	$\Delta \ln \mathrm{D}_{imt}$		$\Delta \ln \mathrm{D}_{imt}$		$\Delta \ln \mathrm{D}_{imt}$	
Time since failure (k)	1 year		2 year		3+ year	
Type of failure	In-Market	Out-of-Market	In-Market	Out-of-Market	In-Market	Out-of-Market
Main Explanatory Variables						
Equity	0.815***	0.815***	0.800***	0.805***	0.797***	0.792***
	(0.058)	(0.059)	(0.058)	(0.058)	(0.058)	(0.059)
NPL	−0.655***	−0.680***	−0.650***	−0.683***	−0.652***	−0.686***
	(0.064)	(0.066)	(0.065)	(0.066)	(0.065)	(0.067)
ROA	2.378***	2.330***	2.290***	2.166***	2.189***	2.056***
	(0.269)	(0.283)	(0.271)	(0.292)	(0.273)	(0.294)
Control Variables						
Implicit Interest Rate	2.522***	2.466***	2.524***	2.532***	2.509***	2.665***
	(0.355)	(0.356)	(0.356)	(0.359)	(0.357)	(0.364)
Ln(Total Assets)	−0.055***	−0.054***	−0.056***	−0.055***	−0.057***	−0.057***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
BHC	−0.019**	−0.019**	−0.019**	−0.017*	−0.018**	−0.016*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Ln(Number of Branches)	−0.010***	−0.010***	−0.011***	−0.012***	−0.012***	−0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HHI	0.028*	0.025	0.032*	0.022	0.036**	0.020
	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)
Cost/Income	0.078***	0.078***	0.075***	0.078***	0.072***	0.075***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)
Real Estate Loans	−0.054***	−0.058***	−0.049***	−0.054***	−0.045***	−0.046***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Liquid Assets	−0.327***	−0.319***	−0.331***	−0.323***	−0.332***	−0.325***
	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Securities for Sale and Investment	−0.185***	−0.183***	−0.186***	−0.186***	−0.186***	−0.183***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Savings Bank	0.139**	0.141**	0.139**	0.138**	0.139**	0.138**
	(0.066)	(0.067)	(0.066)	(0.066)	(0.066)	(0.067)
Interaction Main Explanatory Variables						
F	−0.115***	−0.092***	−0.094**	−0.062**	−0.102***	−0.061**
	(0.039)	(0.030)	(0.038)	(0.028)	(0.036)	(0.027)
$F\times$ Equity	0.272***	0.290***	0.296***	0.210***	0.255***	0.210***
	(0.083)	(0.078)	(0.076)	(0.065)	(0.074)	(0.062)
$F\times$ NPL	−0.521***	−0.274**	−0.398***	−0.126	−0.348***	−0.118
	(0.146)	(0.137)	(0.135)	(0.120)	(0.127)	(0.116)
$F\times$ ROA	0.238	0.838	0.595	1.048**	1.023	1.104**
	(0.760)	(0.608)	(0.703)	(0.530)	(0.663)	(0.509)
Interaction Control Variables						
$F\times$ Implicit Interest Rate	0.614	1.006	0.683	0.997*	0.710	0.534
	(0.673)	(0.650)	(0.658)	(0.585)	(0.628)	(0.570)
$F\times$ Ln(Total Assets)	0.005***	0.002***	0.004***	0.002**	0.004***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$F\times$ BHC	0.010	−0.003	0.005	−0.004	0.005	−0.004
	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
$F\times$ Ln(Number of Branches)	−0.000	0.004**	0.003	0.006***	0.005***	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$F\times$ HHI	0.011	0.016	−0.010	0.028*	−0.017	0.027*
	(0.021)	(0.018)	(0.020)	(0.016)	(0.023)	(0.016)
$F\times$ Cost/Income	0.026	0.036**	0.020	0.012	0.027*	0.008
	(0.019)	(0.017)	(0.017)	(0.016)	(0.016)	(0.015)
$F\times$ Real Estate Loans	−0.038**	−0.017	−0.053***	−0.038***	−0.051***	−0.041***
	(0.017)	(0.013)	(0.016)	(0.014)	(0.017)	(0.014)
$F\times$ Liquid Assets	0.125***	0.018	0.149***	0.067**	0.153***	0.083***
	(0.039)	(0.031)	(0.032)	(0.028)	(0.030)	(0.027)
$F\times$ Securities for Sale and Investment	0.037**	0.012	0.040**	0.031**	0.042**	0.020
	(0.018)	(0.016)	(0.017)	(0.015)	(0.016)	(0.015)
$F\times$ Savings Bank	−0.005	−0.009	−0.008	−0.003	−0.009	−0.008
	(0.012)	(0.015)	(0.010)	(0.012)	(0.010)	(0.010)
Bank Fixed Effects	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included
Within $\bar{R}^2$	0.032	0.031	0.033	0.032	0.034	0.033
N	84,131	84,131	84,131	84,131	84,131	84,131
Failed	0.090	0.107	0.090	0.107	0.090	0.107

This table explores the difference in depositor reaction to bank failures whose headquarters is located in the local market, versus bank failures whose headquarters are located out-of-market. 'Failed' indicates the percentage of markets that experienced a failure during the time period. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

only exert discipline after failures when they have a choice between banks. While it is appealing to interpret these results in this way, we have seen in Table 5.3 that failures occurred less in concentrated markets. The percentage of markets experiencing a failure is much higher for low concentration markets compared to high concentration markets, meaning that this conclusion might be driven by other market characteristics than the degree of concentration. To take this possibility into account, I perform the analysis on MSAs only in Section 5.4.

5.3.3 Is this a general crisis effect?

The results presented here show consistently that failures in local markets lead to a wake-up call of depositors, and that this wake-up call is a long-run phenomenon. However, they still do not rule out that depositor behavior may have changed as a result of the crisis. Berger and Turk-Ariss (2014) indeed find that overall market discipline decreased during the crisis compared to the pre-crisis behavior of depositors. Since the sample used in this chapter coincides with the beginning of the crisis, it is possible that the results presented so far are caused by a general change in depositor behavior due to the crisis. To test this possibility of a wake-up call due to a crisis effect, I extend the dataset to include the pre-crisis years 2002 - 2006. As there are hardly any failures occurring prior to the crisis, a true differences-in-differences model is unfortunately not well-identified, and I therefore estimate the following regression model:

$$\begin{aligned}\Delta \ln D_{i,m,t} &= \beta_0 + \beta_1 \mathbf{Risk}_{i,t-1} + \beta_2 \mathbf{Controls}_{i,m,t-1} \\ &+ \beta_3 C + \beta_4 C\mathbf{Risk}_{i,t-1} + \beta_5 C\mathbf{Controls}_{i,m,t-1} \\ &+ \beta_6 CF_{m,t-k} + \beta_7 CF_{m,t-k}\mathbf{Risk}_{i,t-1} + \beta_8 CF_{m,t-k}\mathbf{Controls}_{i,m,t-1} \\ &+ \beta_i + \beta_m + \beta_t + \epsilon_{i,m,t} \end{aligned} \tag{5.3}$$

where C is a dummy variable taking the value 1 for the years 2007 through 2013 and 0 otherwise, and the rest of the covariates are the same as in Equation (5.2). The coefficients of interest here are the crisis interaction with the vector of risk characteristics, $\beta_4 \times C \times \mathbf{Risk}_{i,t-1}$, and the double interaction capturing the effect of risk when failures occur during the crisis, $\beta_7 \times C \times F_{m,t-k} \times \mathbf{Risk}_{i,t-1}$. The results are presented in

Table 5.7: Do depositors discipline differently in markets with high or low concentration?

	(1) $\Delta \ln D_{imt}$				(2) $\Delta \ln D_{imt}$				(3) $\Delta \ln D_{imt}$			
HHI	HHI $\geq$ 2500				1500 $\leq$ HHI $<$ 2500				HHI $<$ 1500			
Time since failure (k)		1 year	2 year	3+ year		1 year	2 year	3+ year		1 year	2 year	3+ year
Equity	0.752***	0.744***	0.745***	0.737***	0.607***	0.579***	0.576***	0.575***	1.042***	0.985***	0.963***	0.937***
	(0.106)	(0.105)	(0.105)	(0.105)	(0.097)	(0.094)	(0.093)	(0.093)	(0.112)	(0.107)	(0.109)	(0.111)
NPL	−0.570***	−0.518***	−0.540***	−0.556***	−0.587***	−0.561***	−0.590***	−0.623***	−0.877***	−0.850***	−0.859***	−0.853***
	(0.128)	(0.123)	(0.123)	(0.124)	(0.120)	(0.121)	(0.119)	(0.123)	(0.103)	(0.120)	(0.122)	(0.130)
ROA	2.423***	2.196***	2.083***	2.028***	2.384***	2.418***	2.332***	2.163***	2.557***	2.348***	2.031***	1.760***
	(0.521)	(0.526)	(0.527)	(0.530)	(0.462)	(0.492)	(0.502)	(0.511)	(0.462)	(0.545)	(0.588)	(0.611)
F		−0.210***	−0.124*	−0.099		−0.085	−0.057	−0.081		−0.041	−0.034	−0.046
		(0.081)	(0.074)	(0.066)		(0.060)	(0.051)	(0.050)		(0.035)	(0.035)	(0.035)
$F\times$ Equity		0.001	−0.010	0.051		0.397***	0.239**	0.185		0.179**	0.161**	0.201**
		(0.190)	(0.181)	(0.161)		(0.136)	(0.118)	(0.116)		(0.084)	(0.081)	(0.081)
$F\times$ NPL		−0.725	−0.242	−0.041		−0.236	0.044	0.140		−0.063	−0.023	−0.052
		(0.488)	(0.384)	(0.386)		(0.263)	(0.243)	(0.228)		(0.182)	(0.177)	(0.177)
$F\times$ ROA		5.025**	5.313***	4.428**		−0.643	−0.191	0.461		0.460	0.836	1.209
		(2.008)	(1.849)	(1.751)		(1.168)	(1.093)	(1.007)		(0.806)	(0.800)	(0.792)
Control Variables	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Bank Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Within $\bar{R}^2$	0.023	0.025	0.025	0.026	0.024	0.026	0.026	0.026	0.044	0.045	0.046	0.046
N	24,550	24,550	24,550	24,550	29,102	29,102	29,102	29,102	30,479	30,479	30,479	30,479
Failed	0.051	0.051	0.051	0.051	0.102	0.102	0.102	0.102	0.267	0.267	0.267	0.267

This table explores whether depositors react differently across high, moderate or low concentration markets. Markets are divided according to HHI and in accordance with the U.S. Department of Justice. High concentration is defined as HHI values above 2500, moderate concentration is defined as values between 1500 and 2500, and low concentration is defined as values below 1500. Control variables are as indicated in Table 5.1, every specification includes bank, market and time fixed effects to control for demand effects. 'Failed' indicates the percentage of markets that experienced a failure during the time period. Full results are available upon request. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

columns (1)-(4) in Table 5.8.

The table shows that there is a general wake-up effect during the crisis, as banks are disciplined more in the crisis than in the years leading up to it. The positive coefficient on the Crisis$\times$NPL indicates that depositors heavily monitored on this variable prior to the crisis. The total effect of NPL is still negative, which is consistent with market discipline. Disciplining on the other risk variables occurs as well. From the double interaction with the crisis variable *and* the bank failure indicator, we can see that excess discipline is exerted in those markets experiencing failures. Contrary to the results of the baseline specification, the effect of capital and NPL die out 2 years after bank failures. However, similar to the baseline specification, the disciplining effect of Return-on-Assets is present 3 or more years after failure. Overall, the results indicate that, similar to Table 5.4, depositors exert extra discipline in banking markets where failures occurred. The next section explores the robustness of the results presented so far.

5.4 Robustness tests

In this section, I present four robustness tests for the results in the previous section. In the first test, I use the multidimensionality of the dataset to add extra controls for possible unobserved heterogeneity. The second test investigates the choice of MSAs and non-MSA counties as local markets, and explores some alternatives. The third and fourth test address concerns regarding endogeneity, with the third test looking at failures in neighboring markets as a way to control for local business cycle effects and the fourth test excluding banks with a large local presence.

Unobserved heterogeneity The baseline specifications in Tables 5.4 and 5.A.1 include bank, market and year fixed effects to control for unobserved heterogeneity. However, given the dimensions of the dataset, it is possible to add more than the standard fixed effects to control for the heterogeneity. For instance, while the market fixed effects can control for market-specific changes in deposits, and the year fixed effects pick up the countrywide business cycle effects, market$\times$year fixed effects can control for local market business cycle effects. Similarly, bank$\times$market fixed effects can be employed

Table 5.8: Is there a general crisis effect?

Time since failure (k)	(1) $\Delta \ln D_{imt}$	(2) $\Delta \ln D_{imt}$ 1 year	(3) $\Delta \ln D_{imt}$ 2 year	(4) $\Delta \ln D_{imt}$ 3+ year	(5) $\Delta \ln D_{imt}$	(6) $\Delta \ln D_{imt}$ 1 year	(7) $\Delta \ln D_{imt}$ 2 year	(8) $\Delta \ln D_{imt}$ 3+ year	(9) $\Delta \ln D_{imt}$	(10) $\Delta \ln D_{imt}$ 1 year	(11) $\Delta \ln D_{imt}$ 2 year	(12) $\Delta \ln D_{imt}$ 3+ year
Equity	0.183*	0.179*	0.179*	0.180*	0.231**	0.227**	0.227**	0.228**	0.360***	0.350***	0.345***	0.343***
	(0.098)	(0.099)	(0.100)	(0.100)	(0.105)	(0.106)	(0.106)	(0.107)	(0.099)	(0.099)	(0.100)	(0.101)
NPL	−1.877***	−1.853***	−1.848***	−1.848***	−1.798***	−1.780***	−1.777***	−1.777***	−1.902***	−1.891***	−1.895***	−1.900***
	(0.144)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.143)	(0.133)	(0.133)	(0.133)	(0.133)
ROA	0.783	0.766	0.740	0.717	0.586	0.580	0.556	0.522	0.416	0.427	0.396	0.355
	(0.732)	(0.733)	(0.734)	(0.729)	(0.707)	(0.708)	(0.710)	(0.702)	(0.634)	(0.635)	(0.635)	(0.631)
Crisis Interaction												
Crisis	0.373	0.424	0.430	0.458	−0.060	−0.051	−0.049	−0.044	−0.092	−0.140	−0.085	−0.077
	(6.801)	(6.835)	(6.873)	(6.958)	(0.854)	(0.849)	(0.846)	(0.840)	(0.118)	(0.161)	(0.115)	(0.122)
C× Equity	0.324***	0.291***	0.296***	0.299***	0.370***	0.340***	0.347***	0.355***	0.398***	0.368***	0.377***	0.384***
	(0.059)	(0.059)	(0.060)	(0.059)	(0.061)	(0.062)	(0.064)	(0.062)	(0.067)	(0.065)	(0.067)	(0.065)
C× NPL	0.978***	1.025***	0.997***	0.974***	0.966***	1.023***	0.994***	0.972***	1.052***	1.089***	1.039***	1.000***
	(0.270)	(0.275)	(0.264)	(0.258)	(0.280)	(0.288)	(0.278)	(0.275)	(0.203)	(0.211)	(0.203)	(0.199)
C× ROA	3.451***	3.157***	2.983***	2.841***	3.388***	3.147***	2.976***	2.807***	3.823***	3.665***	3.434***	3.168***
	(0.937)	(0.984)	(0.986)	(0.986)	(0.934)	(0.971)	(0.969)	(0.967)	(0.820)	(0.875)	(0.885)	(0.888)
Crisis and Failure Interaction												
$C \times F$		−0.088**	−0.063	−0.062		−0.066	−0.053	−0.065		0.324	−0.034	−0.062
		(0.041)	(0.042)	(0.043)		(0.042)	(0.044)	(0.049)		(1.075)	(0.047)	(0.052)
$C \times F\times$ Equity		0.228***	0.138	0.104		0.245***	0.139	0.074		0.244***	0.138	0.097
		(0.082)	(0.090)	(0.085)		(0.093)	(0.104)	(0.108)		(0.082)	(0.087)	(0.094)
$C \times F\times$ NPL		−0.363**	−0.174	−0.090		−0.354**	−0.152	−0.109		−0.165	0.028	0.107
		(0.163)	(0.143)	(0.152)		(0.174)	(0.156)	(0.181)		(0.178)	(0.154)	(0.180)
$C \times F\times$ ROA		1.670	1.940	2.202*		1.387	1.969	2.386*		0.785	1.519	2.171*
		(1.339)	(1.302)	(1.256)		(1.333)	(1.324)	(1.280)		(1.258)	(1.221)	(1.178)
Control Variables	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Bank Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Bank×Market Fixed Effects					Included	Included	Included	Included	Included	Included	Included	Included
Market×Year Fixed Effects									Included	Included	Included	Included
Within $\bar{R}^2$	0.047	0.048	0.048	0.049	0.055	0.055	0.056	0.056	0.051	0.052	0.052	0.053
N	149,721	149,721	149,721	149,721	149,721	149,721	149,721	149,721	149,721	149,721	149,721	149,721

This table explores the cumulative effect of the crisis and bank failures. Control variables are as indicated in Table 5.1, every specification includes bank, market and time fixed effects to control for demand effects, while columns 5-12 include bank×market fixed effects and columns 9-12 also include market×year fixed effects. Full results are available upon request. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

to control for market-specific strategies, as banks might want to expand their market share in certain markets but not in others. Given the large dimensionality involved in adding bank×market and market×year fixed effects, these estimations are only feasible with the iterative within transformation employed so far. In Table 5.A.3, the extra fixed effects are added, columns (1)-(4) show the result where the baseline specification is augmented with bank×market fixed effects and in columns (5)-(8) both bank×market and market×year fixed effects are added. In the latter specification, variables that vary only over market and time ('HHI' and 'Ln(Number of Branches)') drop out due to perfect collinearity with the added fixed effects. Moreover, the variable indicating whether a failure occurred in the previous year is also dropped for the same reason. This problem does not occur with the dummy variables indicating if failures occurred in markets 2 or 3+ years ago. Overall, the results on the interaction with the risk characteristics are robust to adding these additional fixed effects. Depositors discipline mainly on the capital ratio, and this excess discipline is long-lasting as it is significant even 3+ years after the failure occurred. Similar to the results in Table 5.4, the return-on-assets becomes statistically significant in later years after failure.

I also add the extra fixed effects to regressions in Tables 5.5 and 5.8. Overall, the results are very similar. Specifications (2) and (3) in Table 5.5 show that adding the fixed effects does not change significance of the risk variables in the different stages of the crisis. Moreover, columns (5)-(8) in Table 5.8 show that adding these fixed effects does not change the result when considering a general crisis effect.

Definition of local market The second robustness test examines the definition of local markets. Consistent with previous literature, local markets are defined as MSAs and non-MSA counties. Given the research question at hand, this definition of a local banking market might not be the most appropriate. This chapter tries to identify the effect on depositors who have witnessed a bank failure in their local market, but there is heterogeneity in the size of these markets. For instance, the New York-Newark-Bridgeport MSA consists of 25 counties. If a bank that is only operating branches in a few of these counties fails, the entire MSA is considered to have experienced a failure. While depositors might experience a wake-up call in these counties, it is questionable that this information could impact depositors in other counties belonging to the same

MSA. Nonetheless, in MSAs news might spread faster due to a higher population density and, e.g., a higher share of commuters. Non-MSA counties, on the other hand, may be closer communities where a bank failure can have a stronger impact. Table 5.3 and Section 5.3.2 already provide some preliminary evidence, showing that failures are more likely to occur in larger and less concentrated markets such as MSAs.

In total, the dataset covers banks active in 388 MSAs and 1,940 non-MSA counties. Despite including market fixed effects, the baseline specification only considers the average impact of failures on discipline over both types, and does not allow separate slopes. To test the robustness of this assumption, I first look at counties only, after which I analyze the difference in reactions to failures in MSAs and non-MSA counties separately.

Table 5.A.4 shows the results when defining the county as a local banking market. Columns (1)-(4) display the baseline specification with bank, market and year fixed effects, while columns (5)-(12) add bank$\times$market and market$\times$year fixed effects. The results indicate that general market discipline was present in the crisis, however it seems that in none of the specifications significant excess discipline is being exerted by depositors. While the point estimates have the correct sign and similar magnitude as in Table 5.4, they are not statistically significant and therefore the results from Section 5.3 do not seem robust to the choice of local banking market.

This, of course, warrants a further investigation into the causes and the robustness of the baseline specification. To do so, I revert to the original market definition of MSAs and non-MSA counties. Instead of allowing for a shared slope across both type of markets, like in Table 5.4, I now split up the markets and analyze the effect of failures separately in Table 5.9. The table shows that overall market discipline was present in both MSAs and non-MSA counties. Moreover, the table confirms the previous result that in non-MSA counties (columns (5)-(8)), there was no significant increase in discipline after bank failures. In MSAs (columns (1)-(4)), however, there *was* extra discipline based on the equity ratio of banks, and this effect is long-lasting. These results seem to suggest that failures in non-MSA counties do not lead to a wake-up effect. A possible explanation is that these markets are simply less populated and a lot smaller, and therefore depositors are less exposed to failures. This explanation, however, assumes that MSAs and non-MSA counties are subject to the same amount of bank failures. When looking at the data on bank failures in Table 5.3, it becomes clear this assumption is not correct and that the bulk of the failures occur in MSAs: of the local markets that have seen a failure during the sample, 57% are MSAs, while they make up a much smaller share of the total number of markets. For instance, from the markets that have not experienced a failure, only 14% is a MSA and the remaining 86% are non-MSA counties. The wake-up call thus seems more present in larger banking markets simply because failures have occurred there more often. The results from Section 5.3 are thus still robust, but are driven mainly by the larger markets that have experienced more failures.

Neighboring markets Another concern is that failures do not occur randomly. While Figure 5.2 show that there is variation across markets and states, Table 5.3 and the previous section showed that larger markets suffered more failures. One way to control for this potential endogeneity issue is to look at failures in neighboring markets. These failures are likely due to the same local business cycle effects, but should not impact depositors in the own market as they did not experience them. Moreover, neighboring markets are expected to be similar in both observable and unobservable characteristics (see e.g. Huang, 2008). If experiencing failures leads to extra discipline, I expect to find that failures in neighboring markets would not lead to extra discipline, as depositors in the own market should not be impacted. Table 5.10 shows the estimation output, where F denotes the failure of a bank within the market the bank is active in, while F_n denotes the failure in a neighboring market.

Overall, we can see that the results from the baseline specification hold when adding the neighboring failures. Depositors still react to Equity and NPL, and this is a long-lasting effect. Failures in neighboring markets do not seem to consistently lead to extra discipline in the own market, even though banks on average have a lower growth in deposits if banks in neighboring markets fail. In the specification with only bank, market and year fixed effects, the effects on Equity and NPL of banks in the own market are only significant 3 or more years after the neighboring failure. When adding extra controls in bank$\times$market and market$\times$year fixed effects, it becomes clear that neighboring failures do not consistently lead to extra discipline. However, the effect of Equity after a failure within the own market is always significant and of the correct

Table 5.9: Do failures lead to a wake-up effect in all markets?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$	$\Delta \ln \mathrm{D}_{imt}$
Market	MSAs				Non-MSA Counties			
Time since failure (k)		1 year	2 year	3+ year		1 year	2 year	3+ year
Equity	0.986***	0.897***	0.859***	0.836***	0.714***	0.712***	0.718***	0.717***
	(0.096)	(0.090)	(0.091)	(0.092)	(0.067)	(0.067)	(0.068)	(0.068)
NPL	−0.885***	−0.818***	−0.875***	−0.864***	−0.503***	−0.506***	−0.505***	−0.536***
	(0.103)	(0.122)	(0.124)	(0.132)	(0.074)	(0.076)	(0.076)	(0.077)
ROA	2.431***	2.258***	1.910***	1.666***	2.732***	2.673***	2.622***	2.517***
	(0.406)	(0.489)	(0.521)	(0.545)	(0.346)	(0.350)	(0.353)	(0.355)
F		−0.066**	−0.065**	−0.074**		−0.074	−0.017	−0.029
		(0.032)	(0.031)	(0.032)		(0.061)	(0.048)	(0.042)
$F \times$ Equity		0.278***	0.296***	0.314***		0.097	−0.076	−0.087
		(0.074)	(0.072)	(0.072)		(0.111)	(0.093)	(0.091)
$F \times$ NPL		−0.139	0.010	−0.028		−0.055	0.116	0.308
		(0.147)	(0.147)	(0.150)		(0.265)	(0.215)	(0.198)
$F \times$ ROA		0.234	0.663	0.923		1.953	2.034	2.538**
		(0.669)	(0.671)	(0.677)		(1.656)	(1.267)	(1.080)
Control Variables	Included	Included	Included	Included	Included	Included	Included	Included
Bank Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Within $\bar{R}^2$	0.041	0.042	0.044	0.045	0.026	0.026	0.026	0.026
N	37,203	37,203	37,203	37,203	46,928	46,928	46,928	46,928
Failed	0.274	0.274	0.274	0.274	0.034	0.034	0.034	0.034

This table explores the robustness of the definition of a market, by using using only MSAs or non-MSA counties as the regional market in which depositors move their deposits. Since there were no changes in type between savings and commercial banks in the non-MSA counties, the 'Savings Bank' variable was omitted from all specifications. 'Failed' indicates the percentage of markets that experienced a failure during the time period. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table 5.10: Do neighboring failures lead to a wake-up effect?

	(1)			(2)			(3)		
Time since failure (k)	$\Delta \ln D_{imt}$ 1 year	$\Delta \ln D_{imt}$ 2 year	$\Delta \ln D_{imt}$ 3+ year	$\Delta \ln D_{imt}$ 1 year	$\Delta \ln D_{imt}$ 2 year	$\Delta \ln D_{imt}$ 3+ year	$\Delta \ln D_{imt}$ 1 year	$\Delta \ln D_{imt}$ 2 year	$\Delta \ln D_{imt}$ 3+ year
Equity	0.795***	0.774***	0.753***	0.864***	0.841***	0.812***	0.889***	0.868***	0.843***
	(0.055)	(0.055)	(0.056)	(0.055)	(0.055)	(0.056)	(0.062)	(0.063)	(0.063)
NPL	−0.692***	−0.710***	−0.757***	−0.646***	−0.665***	−0.713***	−0.706***	−0.729***	−0.782***
	(0.065)	(0.067)	(0.070)	(0.064)	(0.067)	(0.071)	(0.075)	(0.078)	(0.083)
ROA	2.186***	1.972***	1.639***	2.033***	1.749***	1.325***	2.621***	2.327***	1.966***
	(0.280)	(0.293)	(0.299)	(0.286)	(0.302)	(0.310)	(0.319)	(0.337)	(0.345)
F	−0.068**	−0.044*	−0.038	−0.048*	−0.035	−0.034		−0.057	−0.080*
	(0.030)	(0.027)	(0.026)	(0.029)	(0.028)	(0.029)		(0.035)	(0.048)
$F\times$ Equity	0.258***	0.190***	0.141**	0.295***	0.220***	0.148*	0.284***	0.220***	0.165*
	(0.071)	(0.064)	(0.063)	(0.077)	(0.077)	(0.084)	(0.078)	(0.080)	(0.088)
$F\times$ NPL	−0.378***	−0.210*	−0.216*	−0.260**	−0.113	−0.131	−0.171	−0.002	0.002
	(0.123)	(0.115)	(0.113)	(0.130)	(0.136)	(0.139)	(0.137)	(0.142)	(0.149)
$F\times$ ROA	0.487	0.801	0.900	0.095	0.411	0.401	−0.104	0.599	0.966
	(0.595)	(0.564)	(0.554)	(0.609)	(0.578)	(0.605)	(0.697)	(0.638)	(0.683)
F_n	−0.028	−0.032*	−0.053***	−0.026	−0.040**	−0.070***		−0.027	−0.058*
	(0.021)	(0.019)	(0.020)	(0.021)	(0.020)	(0.022)		(0.032)	(0.035)
$F_n\times$ Equity	0.022	0.026	0.085*	−0.002	0.013	0.102	−0.025	−0.025	0.051
	(0.054)	(0.049)	(0.049)	(0.057)	(0.054)	(0.062)	(0.062)	(0.058)	(0.065)
$F_n\times$ NPL	0.164	0.101	0.188*	0.064	0.038	0.120	0.116	0.006	0.080
	(0.101)	(0.099)	(0.096)	(0.109)	(0.108)	(0.111)	(0.125)	(0.123)	(0.124)
$F_n\times$ ROA	0.494	0.384	0.699	0.651	0.922*	1.458***	0.055	0.100	0.245
	(0.523)	(0.513)	(0.493)	(0.550)	(0.540)	(0.536)	(0.646)	(0.556)	(0.582)
Control Variables	Included	Included	Included	Included	Included	Included	Included	Included	Included
Bank Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included
Bank×Market Fixed Effects				Included	Included	Included	Included	Included	Included
Market×Year Fixed Effects							Included	Included	Included
Within $\bar{R}^2$	0.032	0.034	0.035	0.037	0.038	0.041	0.038	0.039	0.042
N	84,131	84,131	84,131	84,131	84,131	84,131	84,131	84,131	84,131
Failed	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151
Failed Neighbor	0.198	0.198	0.198	0.198	0.198	0.198	0.198	0.198	0.198

This table explores the effect of bank failures in a neighboring market. F indicates a failure in the home market, while F_n indicates a failure in the neighboring market. Specification (1) includes the standard fixed effects, while specifications (2) and (3) add bank×market and market×year fixed effects, respectively. Since 'HHI' and 'Ln(Number of Branches)' are market-level indicators and vary only over market and time, they are dropped in specification (3). For the same reason, F drops out as it is measures failures in a market during the previous year only, and is therefore perfectly correlated with the market×year fixed effects. 'Failed' indicates the percentage of markets that experienced a failure during the time period, while 'Failed Neighbor' indicates the percentage of neighboring markets experiencing a failure. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

sign. It therefore seems that only experiencing failures in the own market leads to extra depositor discipline, and this channel is robust for local business cycle effects.

Local bank endogeneity A second source of potential endogeneity is the relationship between changes in deposits and a bank's risk characteristics. Even though Equation (5.2) tries to solve this by lagging the vector of bank risk characteristics, there could potentially be a problem if a bank depends on most of its deposit funding from the same market. If banks obtain deposit funding equally from multiple markets, there is no mechanical link between the deposit growth of each bank in a certain market and the bank-level risk characteristic. However, if banks rely mainly on funding from their one single market, there is a potential link between the *bank-market* level change in deposits and *bank* level risk characteristics. While this is partially dealt with in Equation (5.2) by lagging the risk characteristics, the source of the potential endogeneity is removed in order to test the robustness. I classify banks that have at least 75% of their deposits in the market as local banks, and remove these from the sample before redoing the analysis. Table 5.11 shows the results, which are similar to the baseline specification: extra discipline is mainly directed at the capital ratio of banks and the share of non-performing loans the first year after bank failure. In the years afterwards, discipline is again mainly directed at the capital ratio of banks, although similar to the baseline specification the return-on-assets also becomes statistically significant.

5.5 Conclusion

This chapter has looked at whether depositor discipline was present during the crisis, and whether the bank failures have indeed increased discipline by serving as a 'wake-up call'. In doing so, this chapter offers two contributions compared to the existing literature. First, instead of measuring discipline at the level of the bank entity, this chapter looks at depositor discipline in local markets. The focus on the local market is motivated by the fact that depositors generally only choose to keep their savings at banks that are active in their vicinity. By comparing banks at a national or even supranational level, this fact is not taken into account and therefore offers a distorted

Table 5.11: Do failures lead to a wake-up effect when removing local banks?

	(1)	(2)	(3)	(4)
	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$
Time since failure (k)		1 year	2 year	3+ year
Equity	0.448***	0.428***	0.425***	0.424***
	(0.059)	(0.059)	(0.059)	(0.060)
NPL	−0.616***	−0.572***	−0.596***	−0.614***
	(0.075)	(0.078)	(0.078)	(0.080)
ROA	3.010***	2.909***	2.736***	2.557***
	(0.310)	(0.329)	(0.337)	(0.342)
F		−0.062*	−0.047	−0.062**
		(0.036)	(0.032)	(0.031)
$F\times$ Equity		0.163**	0.118*	0.109*
		(0.072)	(0.064)	(0.062)
$F\times$ NPL		−0.266*	−0.046	−0.022
		(0.157)	(0.142)	(0.134)
$F\times$ ROA		0.888	1.117*	1.403**
		(0.708)	(0.646)	(0.616)
Control Variables	Included	Included	Included	Included
Bank Fixed Effects	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included
Within $\bar{R}^2$	0.016	0.017	0.018	0.018
N	63,440	63,440	63,440	63,440
Failed	0.125	0.125	0.125	0.125

This table explores the robustness of the results when removing banks that are considered local (i.e. that have at least 75% of their deposits in the market). As the changes in deposits might lead to a worse risk profile in the next period, these banks are removed for possible endogeneity concerns. 'Failed' indicates the percentage of markets that experienced a failure during the time period. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

identification of depositor discipline. Second, by analyzing the behavior of depositors in local markets, this chapter presents a cleaner identification of the wake-up call than previous studies have managed, as not all local markets have seen bank failures and therefore experienced the potential for a wake-up call.

I find that discipline was present in local markets between 2007 and 2013, and that depositors mainly focused on the capital buffers of banks. Furthermore, discipline was more severe in those local markets that saw a failure, and the analyses show that this is a long-run phenomenon. Depositors were also able to distinguish between different types of failures. Whereas they reacted more heavily to failures of banks with main offices within their local market, out-of-market and out-of-state failures did not give depositors the incentive to increase monitoring on the remaining banks. Finally, since more failures occurred in larger banking markets, the effect seems to be mainly driven by these markets.

The findings presented in this chapter have important implications for the regulatory and supervisory set-up following the crisis. Despite interventions such as bailouts and increases in the level of deposit insurance, depositors still seem to monitor bank riskiness and adjust their deposits holdings accordingly. Moreover, depositor discipline is more present in markets in which the FDIC has let banks fail, even though the depositors were often fully compensated. These failures, possibly serving as a reminder regarding banks inherent riskiness, can actually lead to more involved depositors and potentially a safer banking system. Finally, the Net Stable Funding Ratio (NFSR) introduced in Basel III will force banks to hold a sufficient amount of stable funding, which includes demand and other customer deposits. This chapter shows that deposit funding of banks depends not only on the bank's risk characteristics, but also on what is happening to other banks in their market.

Appendix 5.A Additional figures and tables

Table 5.A.1: Do failures lead to a wake-up effect?

	(1)	(2)	(3)	(4)
	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$
Time since failure (k)		1 year	2 year	3+ year
Main Explanatory Variables				
Equity	0.853***	0.801***	0.790***	0.784***
	(0.063)	(0.058)	(0.058)	(0.059)
NPL	−0.725***	−0.657***	−0.670***	−0.683***
	(0.064)	(0.067)	(0.067)	(0.068)
ROA	2.435***	2.277***	2.091***	1.920***
	(0.264)	(0.286)	(0.293)	(0.298)
Control Variables				
Implicit Interest Rate	2.347***	2.580***	2.629***	2.709***
	(0.346)	(0.360)	(0.361)	(0.362)
Ln(Total Assets)	−0.053***	−0.055***	−0.056***	−0.058***
	(0.004)	(0.005)	(0.005)	(0.005)
BHC	−0.020**	−0.018**	−0.017*	−0.016*
	(0.009)	(0.009)	(0.009)	(0.009)
Ln(Number of Branches)	−0.009***	−0.011***	−0.012***	−0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
HHI	0.026	0.027	0.027	0.028
	(0.017)	(0.017)	(0.018)	(0.018)
Cost/Income	0.083***	0.077***	0.076***	0.071***
	(0.009)	(0.009)	(0.009)	(0.010)
Real Estate Loans	−0.057***	−0.055***	−0.050***	−0.043***
	(0.014)	(0.014)	(0.014)	(0.014)
Liquid Assets	−0.319***	−0.324***	−0.329***	−0.330***
	(0.017)	(0.018)	(0.018)	(0.019)
Securities for Sale and Investment	−0.181***	−0.184***	−0.187***	−0.185***
	(0.012)	(0.012)	(0.012)	(0.012)
Savings Bank	0.140**	0.140**	0.137**	0.139**
	(0.067)	(0.067)	(0.066)	(0.068)
Interaction Main Explanatory Variables				
F		−0.086***	−0.066***	−0.076***
		(0.028)	(0.026)	(0.025)
$F\times$ Equity		0.266***	0.206***	0.196***
		(0.065)	(0.060)	(0.058)
$F\times$ NPL		−0.299**	−0.166	−0.123
		(0.120)	(0.111)	(0.108)
$F\times$ ROA		0.846	1.142**	1.433***
		(0.550)	(0.510)	(0.488)
Interaction Control Variables				
$F\times$ Implicit Interest Rate		0.895	1.068**	0.994**
		(0.548)	(0.520)	(0.489)
$F\times$ Ln(Total Assets)		0.003***	0.002***	0.003***
		(0.001)	(0.001)	(0.001)
$F\times$ BHC		−0.000	0.001	0.000
		(0.006)	(0.006)	(0.006)
$F\times$ Ln(Number of Branches)		0.004**	0.005***	0.006***
		(0.002)	(0.001)	(0.001)
$F\times$ HHI		0.008	0.008	0.004
		(0.015)	(0.013)	(0.015)
$F\times$ Cost/Income		0.027*	0.009	0.013
		(0.015)	(0.014)	(0.014)
$F\times$ Real Estate Loans		−0.027**	−0.040***	−0.041***
		(0.013)	(0.013)	(0.012)
$F\times$ Liquid Assets		0.055*	0.088***	0.095***
		(0.029)	(0.027)	(0.026)
$F\times$ Securities for Sale and Investment		0.015	0.027**	0.023*
		(0.014)	(0.014)	(0.014)
$F\times$ Savings Bank		−0.008	−0.007	−0.014
		(0.011)	(0.009)	(0.008)
Bank Fixed Effects	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included
Within $\bar{R}^2$	0.031	0.032	0.033	0.034
N	84,131	84,131	84,131	84,131

This table shows the full estimation results for Equation (5.2), including control variables that were left out in Table 5.4 for reason of brevity. Similar to Table 5.4, every specification includes bank, market and time fixed effects to control for demand effects. Column 2 measures the impact on market discipline 1 year after the failure has occurred, column 3 measures the long-run effect 2 years after failure and column 4 measures the long-run effect 3 or more years after bank failure. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table 5.A.2: Is there a difference between in-state and out-of-state failures?

	(1)		(2)		(3)	
	$\Delta \ln D_{imt}$		$\Delta \ln D_{imt}$		$\Delta \ln D_{imt}$	
Time since failure (k)	1 year		2 year		3+ year	
Type of failure	In-State	Out-of-State	In-State	Out-of-State	In-State	Out-of-State
Main Explanatory Variables						
Equity	0.812***	0.814***	0.791***	0.804***	0.786***	0.794***
	(0.058)	(0.059)	(0.058)	(0.058)	(0.058)	(0.058)
NPL	−0.676***	−0.668***	−0.666***	−0.656***	−0.667***	−0.675***
	(0.065)	(0.066)	(0.066)	(0.065)	(0.066)	(0.066)
ROA	2.455***	2.325***	2.354***	2.144***	2.210***	2.016***
	(0.275)	(0.279)	(0.277)	(0.289)	(0.281)	(0.292)
Control Variables						
Implicit Interest Rate	2.466***	2.526***	2.525***	2.559***	2.566***	2.658***
	(0.354)	(0.357)	(0.357)	(0.361)	(0.360)	(0.361)
Ln(Total Assets)	−0.054***	−0.054***	−0.055***	−0.056***	−0.056***	−0.057***
	(0.005)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
BHC	−0.020**	−0.019**	−0.019**	−0.017*	−0.018**	−0.016*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Ln(Number of Branches)	−0.010***	−0.011***	−0.010***	−0.012***	−0.011***	−0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
HHI	0.028	0.026	0.029*	0.025	0.033*	0.026
	(0.017)	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)
Cost/Income	0.080***	0.078***	0.077***	0.077***	0.073***	0.073***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Real Estate Loans	−0.058***	−0.056***	−0.053***	−0.052***	−0.047***	−0.045***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Liquid Assets	−0.327***	−0.319***	−0.332***	−0.322***	−0.333***	−0.322***
	(0.017)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
Securities for Sale and Investment	−0.185***	−0.185***	−0.186***	−0.185***	−0.185***	−0.184***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
Savings Bank	0.139**	0.140**	0.136**	0.138**	0.137**	0.138**
	(0.066)	(0.067)	(0.065)	(0.066)	(0.066)	(0.067)
Interaction Main Explanatory Variables						
F	−0.129***	−0.078***	−0.100***	−0.062**	−0.106***	−0.070**
	(0.034)	(0.030)	(0.033)	(0.028)	(0.030)	(0.028)
$F\times$ Equity	0.380***	0.240***	0.360***	0.200***	0.320***	0.201***
	(0.078)	(0.074)	(0.072)	(0.067)	(0.067)	(0.064)
$F\times$ NPL	−0.368***	−0.316**	−0.286**	−0.230*	−0.254**	−0.137
	(0.143)	(0.130)	(0.121)	(0.126)	(0.115)	(0.124)
$F\times$ ROA	−0.115	0.862	0.072	1.270**	0.530	1.466***
	(0.660)	(0.636)	(0.626)	(0.567)	(0.587)	(0.539)
Interaction Control Variables						
$F\times$ Implicit Interest Rate	0.826	0.783	0.596	1.042*	0.509	0.831
	(0.569)	(0.614)	(0.573)	(0.601)	(0.545)	(0.571)
$F\times$ Ln(Total Assets)	0.004***	0.002***	0.003***	0.002**	0.004***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$F\times$ BHC	0.008	0.001	0.007	−0.004	0.005	−0.003
	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)
$F\times$ Ln(Number of Branches)	0.001	0.004**	0.001	0.007***	0.003**	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$F\times$ HHI	0.004	0.018	0.001	0.025	−0.010	0.016
	(0.020)	(0.018)	(0.015)	(0.017)	(0.018)	(0.016)
$F\times$ Cost/Income	0.016	0.035**	0.002	0.019	0.011	0.018
	(0.018)	(0.017)	(0.017)	(0.016)	(0.015)	(0.015)
$F\times$ Real Estate Loans	−0.003	−0.037**	−0.025*	−0.048***	−0.028*	−0.050***
	(0.015)	(0.014)	(0.014)	(0.015)	(0.015)	(0.014)
$F\times$ Liquid Assets	0.109***	0.023	0.128***	0.068**	0.133***	0.081***
	(0.034)	(0.035)	(0.030)	(0.031)	(0.028)	(0.029)
$F\times$ Securities for Sale and Investment	0.030*	0.022	0.035**	0.031**	0.027*	0.028*
	(0.016)	(0.017)	(0.015)	(0.015)	(0.015)	(0.016)
$F\times$ Savings Bank	−0.007	−0.008	−0.011	−0.002	−0.017**	−0.010
	(0.011)	(0.013)	(0.009)	(0.010)	(0.009)	(0.009)
Bank Fixed Effects	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included
Within $\bar{R}^2$	0.032	0.032	0.033	0.033	0.033	0.033
N	84,131	84,131	84,131	84,131	84,131	84,131
Failed	0.094	0.110	0.094	0.110	0.094	0.110

This table explores the difference in depositor reaction to bank failures whose headquarters is located in the same state, versus bank failures whose headquarters are located out-of-state. 'Failed' indicates the percentage of markets that experienced a failure during the time period. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table 5.A.3: Do failures lead to a wake-up effect?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$	$\Delta \ln D_{imt}$
Time since failure (k)		1 year	2 year	3+ year		1 year	2 year	3+ year
Main Explanatory Variables								
Equity	0.916***	0.868***	0.858***	0.850***	0.943***	0.891***	0.876***	0.869***
	(0.063)	(0.058)	(0.058)	(0.058)	(0.070)	(0.066)	(0.066)	(0.066)
NPL	−0.681***	−0.624***	−0.634***	−0.650***	−0.718***	−0.681***	−0.707***	−0.735***
	(0.064)	(0.066)	(0.067)	(0.069)	(0.073)	(0.077)	(0.078)	(0.081)
ROA	2.268***	2.148***	1.977***	1.795***	2.668***	2.671***	2.442***	2.210***
	(0.270)	(0.294)	(0.303)	(0.309)	(0.299)	(0.327)	(0.340)	(0.348)
Control Variables								
Implicit Interest Rate	2.211***	2.155***	2.194***	2.334***	1.939***	1.979***	2.124***	2.295***
	(0.346)	(0.354)	(0.357)	(0.361)	(0.392)	(0.403)	(0.409)	(0.412)
Ln(Total Assets)	−0.044***	−0.044***	−0.046***	−0.047***	−0.055***	−0.055***	−0.056***	−0.058***
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
BHC	−0.021**	−0.022**	−0.021**	−0.020**	−0.019**	−0.019**	−0.019**	−0.018*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
Ln(Number of Branches)	−0.039***	−0.038***	−0.038***	−0.038***				
	(0.005)	(0.005)	(0.005)	(0.005)				
HHI	0.032*	0.031*	0.032*	0.034*				
	(0.017)	(0.017)	(0.018)	(0.018)				
Cost/Income	0.085***	0.081***	0.082***	0.076***	0.079***	0.079***	0.078***	0.068***
	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)
Real Estate Loans	−0.059***	−0.060***	−0.058***	−0.054***	−0.048***	−0.052***	−0.054***	−0.052***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)	(0.016)
Liquid Assets	−0.319***	−0.317***	−0.326***	−0.338***	−0.316***	−0.317***	−0.330***	−0.345***
	(0.017)	(0.018)	(0.018)	(0.019)	(0.019)	(0.020)	(0.020)	(0.021)
Securities for Sale and Investment	−0.180***	−0.183***	−0.187***	−0.187***	−0.190***	−0.197***	−0.207***	−0.207***
	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)
Savings Bank	0.115*	0.116*	0.114*	0.116*	0.153**	0.153**	0.151**	0.153**
	(0.060)	(0.060)	(0.060)	(0.061)	(0.075)	(0.075)	(0.075)	(0.077)
Interaction Main Explanatory Variables								
F		−0.066**	−0.063**	−0.090***			−0.061	−0.101*
		(0.028)	(0.026)	(0.026)			(0.037)	(0.053)
$F\times$ Equity		0.291***	0.231***	0.220***		0.267***	0.204***	0.198***
		(0.070)	(0.069)	(0.068)		(0.073)	(0.072)	(0.071)
$F\times$ NPL		−0.231*	−0.105	−0.083		−0.112	−0.008	0.031
		(0.128)	(0.130)	(0.131)		(0.133)	(0.136)	(0.141)
$F\times$ ROA		0.562	1.121**	1.488***		−0.067	0.682	1.135*
		(0.557)	(0.522)	(0.519)		(0.615)	(0.590)	(0.586)
Interaction Control Variables								
$F\times$ Implicit Interest Rate		0.310	0.681	0.549		−2.390***	−2.972***	−3.785***
		(0.508)	(0.508)	(0.491)		(0.766)	(0.765)	(0.769)
$F\times$ Ln(Total Assets)		0.000	0.001	0.002***		0.000	0.001*	0.003***
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)
$F\times$ BHC		0.005	0.006	0.005		0.001	0.003	0.003
		(0.006)	(0.006)	(0.006)		(0.006)	(0.006)	(0.006)
$F\times$ Ln(Number of Branches)		0.001	0.003*	0.004**				
		(0.002)	(0.002)	(0.002)				
$F\times$ HHI		−0.000	0.001	−0.004				
		(0.015)	(0.013)	(0.016)				
$F\times$ Cost/Income		0.011	0.000	0.013		0.001	0.001	0.021
		(0.016)	(0.015)	(0.015)		(0.016)	(0.016)	(0.016)
$F\times$ Real Estate Loans		0.013	−0.004	−0.008		0.031**	0.024	0.024
		(0.013)	(0.014)	(0.014)		(0.015)	(0.016)	(0.017)
$F\times$ Liquid Assets		0.003	0.061**	0.109***		−0.004	0.056*	0.107***
		(0.032)	(0.030)	(0.030)		(0.034)	(0.031)	(0.032)
$F\times$ Securities for Sale and Investment		0.019	0.039**	0.036**		0.038**	0.064***	0.058***
		(0.017)	(0.017)	(0.017)		(0.019)	(0.018)	(0.018)
$F\times$ Savings Bank		−0.004	−0.004	−0.015		0.001	0.003	−0.006
		(0.013)	(0.010)	(0.010)		(0.014)	(0.011)	(0.012)
Bank Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Bank×Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included
Market×Year Fixed Effects					Included	Included	Included	Included
Within $\bar{R}^2$	0.036	0.036	0.037	0.037	0.036	0.037	0.038	0.039
N	84,131	84,131	84,131	84,131	84,131	84,131	84,131	84,131

This table shows robustness analyses for the estimation results of Equation (5.2). Compared to Table 5.4, this table exploits the multi-dimensionality of the dataset by adding bank×market fixed effects in columns 1-8 and market×year fixed effects in columns 5-8. The bank×market fixed effects control for different depositor reaction in each bank-market pairing, and the market×year fixed effects control for a different depositor reaction per market over time. Since 'HHI' and 'Ln(Number of Branches)' are market-level indicators and vary only over market and time, they are dropped in columns 5-8. For the same reason, F drops out in column 5, as it is measures failures in a market during the previous year only, and is therefore perfectly correlated with the market×year fixed effects. The results in Table 5.4 are robust to adding these control variables. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

Table 5.A.4: Do failures lead to a wake-up effect at the county level?

Time since failure (k)	(1) $\Delta \ln D_{imt}$	(2) $\Delta \ln D_{imt}$ 1 year	(3) $\Delta \ln D_{imt}$ 2 year	(4) $\Delta \ln D_{imt}$ 3+ year	(5) $\Delta \ln D_{imt}$	(6) $\Delta \ln D_{imt}$ 1 year	(7) $\Delta \ln D_{imt}$ 2 year	(8) $\Delta \ln D_{imt}$ 3+ year	(9) $\Delta \ln D_{imt}$	(10) $\Delta \ln D_{imt}$ 1 year	(11) $\Delta \ln D_{imt}$ 2 year	(12) $\Delta \ln D_{imt}$ 3+ year
Main Explanatory Variables												
Equity	0.660***	0.636***	0.619***	0.610***	0.728***	0.710***	0.699***	0.681***	0.704***	0.685***	0.672***	0.649***
	(0.161)	(0.160)	(0.160)	(0.159)	(0.165)	(0.165)	(0.168)	(0.167)	(0.168)	(0.166)	(0.168)	(0.164)
NPL	−0.843***	−0.803***	−0.800***	−0.820***	−0.792***	−0.759***	−0.769***	−0.787***	−0.795***	−0.772***	−0.791***	−0.801***
	(0.128)	(0.136)	(0.132)	(0.129)	(0.125)	(0.134)	(0.133)	(0.135)	(0.118)	(0.128)	(0.129)	(0.134)
ROA	2.200**	2.055**	1.940**	1.713*	2.010**	1.903**	1.771*	1.523	2.703***	2.747***	2.562**	2.299**
	(0.893)	(0.961)	(0.989)	(0.988)	(0.836)	(0.918)	(0.956)	(0.940)	(0.831)	(0.976)	(1.016)	(1.011)
Control Variables												
Implicit Interest Rate	0.475	0.649	0.831	1.031	0.407	0.332	0.494	0.753	−0.041	−0.061	0.259	0.571
	(1.427)	(1.409)	(1.368)	(1.336)	(1.459)	(1.426)	(1.364)	(1.314)	(1.269)	(1.254)	(1.213)	(1.181)
Ln(Total Assets)	−0.046**	−0.048***	−0.049***	−0.052***	−0.035	−0.035	−0.037*	−0.040*	−0.042**	−0.042**	−0.043***	−0.047***
	(0.018)	(0.018)	(0.017)	(0.017)	(0.022)	(0.022)	(0.022)	(0.021)	(0.017)	(0.017)	(0.016)	(0.015)
BHC	−0.027**	−0.026**	−0.025**	−0.024**	−0.033***	−0.033***	−0.032***	−0.031***	−0.034***	−0.034***	−0.033***	−0.032***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Ln(Number of Branches)	−0.015***	−0.016***	−0.017***	−0.018***	−0.046***	−0.046***	−0.046***	−0.046***				
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)	(0.007)	(0.007)	(0.007)				
HHI	0.027*	0.025	0.024	0.025	0.033*	0.030	0.029*	0.030*				
	(0.016)	(0.016)	(0.016)	(0.016)	(0.019)	(0.018)	(0.017)	(0.018)				
Cost/Income	0.096***	0.094***	0.094***	0.087***	0.098***	0.097***	0.100***	0.093***	0.099***	0.100***	0.100***	0.091***
	(0.031)	(0.029)	(0.028)	(0.027)	(0.032)	(0.029)	(0.029)	(0.026)	(0.032)	(0.030)	(0.030)	(0.028)
Real Estate Loans	−0.055	−0.054	−0.049	−0.040	−0.059	−0.058	−0.056	−0.050	−0.046	−0.048	−0.048	−0.045
	(0.048)	(0.047)	(0.047)	(0.047)	(0.049)	(0.048)	(0.048)	(0.049)	(0.043)	(0.044)	(0.044)	(0.044)
Liquid Assets	−0.330***	−0.327***	−0.329***	−0.328***	−0.334***	−0.322***	−0.323***	−0.329***	−0.333***	−0.324***	−0.328***	−0.334***
	(0.045)	(0.047)	(0.047)	(0.049)	(0.046)	(0.047)	(0.047)	(0.051)	(0.046)	(0.048)	(0.048)	(0.053)
Securities for Sale and Investment	−0.200***	−0.197***	−0.198***	−0.195***	−0.197***	−0.194***	−0.197***	−0.197***	−0.207***	−0.208***	−0.214***	−0.216***
	(0.036)	(0.036)	(0.037)	(0.038)	(0.036)	(0.037)	(0.037)	(0.039)	(0.040)	(0.040)	(0.040)	(0.041)
Savings Bank	0.098	0.098	0.095	0.094	0.103**	0.104**	0.102**	0.101*	0.111**	0.113**	0.111**	0.110**
	(0.070)	(0.070)	(0.072)	(0.074)	(0.050)	(0.050)	(0.051)	(0.053)	(0.046)	(0.046)	(0.046)	(0.047)
Interaction Main Explanatory Variables												
F		−0.065	−0.049	−0.064		−0.038	−0.025	−0.062			−0.047	−0.108
		(0.047)	(0.047)	(0.049)		(0.053)	(0.055)	(0.061)			(0.059)	(0.072)
$F\times$ Equity		0.110	0.118	0.118		0.127	0.086	0.100		0.116	0.061	0.102
		(0.109)	(0.114)	(0.109)		(0.137)	(0.154)	(0.164)		(0.131)	(0.149)	(0.168)
$F\times$ NPL		−0.190	−0.114	−0.052		−0.141	−0.008	0.009		−0.132	−0.041	−0.059
		(0.158)	(0.142)	(0.153)		(0.184)	(0.175)	(0.204)		(0.185)	(0.177)	(0.219)
$F\times$ ROA		1.146	1.172	1.507		0.731	1.219	1.708		−0.138	0.575	1.098
		(1.413)	(1.348)	(1.339)		(1.528)	(1.527)	(1.519)		(1.515)	(1.508)	(1.487)
Interaction Control Variables												
$F\times$ Implicit Interest Rate		1.760**	1.628**	1.286*		0.875	0.812	0.474		−0.365	−1.403	−2.417
		(0.796)	(0.773)	(0.777)		(0.868)	(0.869)	(0.876)		(1.487)	(1.667)	(1.939)
$F\times$ Ln(Total Assets)		0.002***	0.002**	0.003***		0.000	0.000	0.002		0.000	0.002*	0.004**
		(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.002)		(0.001)	(0.001)	(0.002)
$F\times$ BHC		0.003	0.003	0.002		0.010	0.008	0.008		0.004	0.002	0.003
		(0.007)	(0.007)	(0.007)		(0.007)	(0.008)	(0.008)		(0.007)	(0.007)	(0.008)
$F\times$ Ln(Number of Branches)		0.003*	0.005**	0.006***		0.002	0.004	0.005*				
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.003)				
$F\times$ HHI		0.021	0.019	0.014		0.019	0.015	0.012				
		(0.017)	(0.017)	(0.017)		(0.018)	(0.019)	(0.019)				
$F\times$ Cost/Income		0.023	0.005	0.010		−0.001	−0.011	0.002		−0.010	−0.003	0.016
		(0.033)	(0.029)	(0.030)		(0.033)	(0.032)	(0.037)		(0.032)	(0.030)	(0.033)
$F\times$ Real Estate Loans		−0.026	−0.044	−0.046		0.012	−0.015	−0.020		0.027	0.015	0.020
		(0.029)	(0.028)	(0.028)		(0.033)	(0.035)	(0.038)		(0.036)	(0.038)	(0.042)
$F\times$ Liquid Assets		0.012	0.045	0.057		−0.060	−0.017	0.031		−0.058	−0.018	0.028
		(0.045)	(0.044)	(0.045)		(0.051)	(0.048)	(0.061)		(0.051)	(0.047)	(0.060)
$F\times$ Securities for Sale and Investment		−0.026	−0.012	−0.013		−0.021	0.002	0.005		0.007	0.023	0.026
		(0.022)	(0.021)	(0.023)		(0.027)	(0.026)	(0.030)		(0.031)	(0.029)	(0.031)
$F\times$ Savings Bank		−0.006	0.001	−0.003		0.001	0.007	0.003		0.004	0.010	0.009
		(0.012)	(0.010)	(0.011)		(0.015)	(0.012)	(0.014)		(0.017)	(0.014)	(0.015)
Bank Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Market Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Year Fixed Effects	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Bank×Market Fixed Effects					Included	Included	Included	Included	Included	Included	Included	Included
Market×Year Fixed Effects									Included	Included	Included	Included
Within $\bar{R}^2$	0.026	0.027	0.028	0.028	0.028	0.028	0.029	0.029	0.026	0.026	0.027	0.028
N	109,223	109,223	109,223	109,223	109,223	109,223	109,223	109,223	109,223	109,223	109,223	109,223
Failed	0.128	0.128	0.128	0.128	0.128	0.128	0.128	0.128	0.128	0.128	0.128	0.128

This table explores the robustness of the definition of a market, by using a county as the regional market in which depositors move their deposits. Bank×market fixed effects are added in columns 5-12, and market×year fixed effects are added in columns 9-12. The bank×market fixed effects control for different depositor reaction in each bank-market pairing, and the market×year fixed effects control for a different depositor reaction per market over time. Since 'HHI' and 'Ln(Number of Branches)' are market-level indicators and vary only over market and time, they are dropped in columns 9-12. For the same reason, F drops out in column 9, as it is measures failures in a market during the previous year only, and is therefore perfectly correlated with the market×year fixed effects. Standard errors in parentheses, clustered on market. * significant at 10 percent; ** significant at 5 percent; *** significant at 1 percent.

6

Carrying the (Paper) Burden: A Portfolio View of Systemic Risk and Optimal Bank Size*

6.1 Introduction

As a result of the financial crisis, the health and safety of the financial system is at the heart of many policy agendas. Concerns regarding the financial system tend to relate mostly to commercial banks and their parent holding companies. Policy discussions focus either on the riskiness of individual financial institutions, or on what is broadly termed systemic risk. Regarding individual banks, the key question debated is whether some banks are too big: too big to fail, too big with respect to their country's GDP (Bertay et al., 2013; Demirgüç-Kunt and Huizinga, 2013), too big to produce at minimum average costs (see e.g. Wheelock and Wilson, 2012; Hughes and Mester, 2013), or even too big to rescue. Systemic risk discussions are much broader, and may consider the stability of the financial system itself, the macro effects of a shock to that

*This chapter is based on joint work with Jaap W.B. Bos (Maastricht University) and Victoria Purice (Ghent University).

system, or the optimal supervisory setup for dealing with and minimizing the likelihood of such a shock.

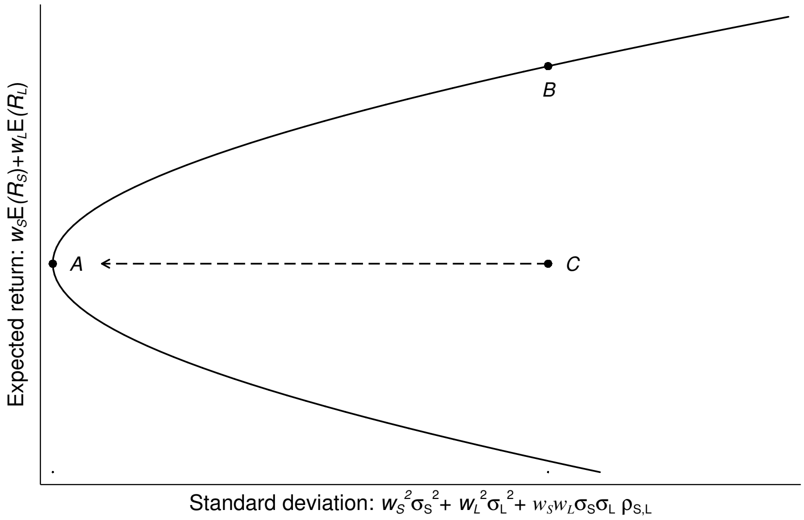
In this chapter, we combine these two discussions and investigate whether the size of the largest banks in the system has contributed to an increase in systemic risk. We do so by engaging the reader in a thought experiment. We imagine a bank supervisor as an investor holding a portfolio of banks. Each bank aims to maximize profits, but thereby incurs a certain amount of risk. Given that banks' profits are not all perfectly correlated, the risk-return relationship of the portfolio that the supervisor holds is expected to be better than that of the riskiest banks in the system on their own. Taking the long-term view, the bank supervisor not only wants to minimize risk but is certainly also interested in return, as high charter values may boost the stability of individual banks.

Although we consider our view of the bank supervisor a thought experiment, recent events have shown that its experimental nature is closer to the reality of a crisis than one may at first suspect. In theory, the bank supervisor mainly represents the interests of deposit holders and deposit insurance guarantees those interests to a large extent. However, during the recent crisis, most supervisors went above and beyond that objective. In the U.S., the Trouble Asset Relief Program (TARP) initially provided support in terms of bank equity share purchases valued at more than three times the total amounts of deposits in the system, although much of these funds were later reclaimed as shares were sold in the market. Moreover, many assets were purchased well above their actual value, resulting in an implicit subsidy of the banking sector (Office of the Special Inspector General for the Troubled Asset Relief Program, 2013). Finally, the Safe, Accountable, Fair & Efficient (SAFE) Banking Act proposed in 2012 gives regulators additional powers to limit bank size in order to lower systemic risk.

Nevertheless, unlike the typical investor, the bank supervisor is seriously limited in buying and selling assets in order to reach or remain at the optimal frontier as depicted in Figure 6.1. As the crisis has shown, even this highly constrained investor can rebalance the weights of the banks in the portfolio, through orderly liquidation and other interventions by the Financial Stability Board such as the capital surcharge for Systemically Important Financial Institutions (SIFIs).

Using this scenario, we pose three questions, each related to the situation depicted

Figure 6.1: The supervisory view: Markowitz efficient frontier, systemic risk and bank size



in Figure 6.1. First, we ask whether large banks offer attractive investment opportunities for the bank supervisor or in other words, whether large banks are characterized by a risk-return relationship superior to that of the other banks in the system. This should establish whether the inclusion of large banks has brought the supervisory portfolio closer to the efficient frontier, e.g., moving from C to B in Figure 6.1.

Second, we examine what would happen to the portfolio's return if the bank supervisor held the minimum variance portfolio. In light of the example in Figure 6.1, would the return move from B to A, thereby requiring the supervisor to give up return in order to hold a less risky portfolio, or would it move from C to A, allowing the choice of a less risky portfolio without sacrificing returns?

Third, we examine whether the bank supervisor should reduce investments in large banks in order to achieve the minimum variance portfolio. In light of the example in Figure 6.1, we ask whether the supervisor has to increase or reduce the share of large banks (s_L) in the portfolio by moving to A. We also examine whether the differences in risk and return between the original portfolio held by the bank supervisor and the new minimum variance portfolio merely reflect a change in the weights of banks, or whether they are driven by high correlation of the returns of the largest banks in the original portfolio.

To perform our experiment, we examine developments in the U.S. banking market since 1984. Using quarterly data on banks' assets and profits, we construct two types of bank supervisory portfolios. For each Federal Reserve District, we include all unconsolidated commercial banks located within the district. For the U.S. as a whole, we construct a portfolio comprising all Bank Holding Companies. In both cases, we are interested in the design of the minimum variance portfolio and how it compares with the actual portfolio.

Our findings indicate that the current portfolios are not located on the efficient frontier, as risk can be reduced without sacrificing return in order to attain the minimum variance portfolio. Moreover, we find that the largest banks in the Federal Reserve portfolios consistently have a significantly lower weight in the hypothetical minimum variance portfolio. In addition, the minimum variance portfolio does not allow for large levels of concentration in the first place, with its weights being much more evenly spread across banks. These results hold even after relaxing some assumptions and allowing the correlation structure to change with the size of banks. In obtaining the minimum variance portfolio, we assume that the supervisor is able to alter the relative weights of the individual banks in the portfolio, along the lines of the Safe, Accountable, Fair & Efficient (SAFE) Banking Act proposed in 2012. The Act was designed to make banks small enough to fail without causing global panic, using regulatory caps. We argue that our results provide important insights into the optimal design of a portfolio of banks, held by a risk-averse supervisor who prefers to incur the least possible amount of risk. The findings suggest reducing the size of the largest banks in the financial system may not only make individual banks safer and easier to fail or rescue, but can also contribute to a reduction in the riskiness of the system as a whole.

This chapter proceeds as follows. In Section 6.2 we discuss our methodology and data, followed by Section 6.3 which contains our results. Extensions and policy implications are addressed in Section 6.4, after which we conclude.

6.2 Data and methodology

6.2.1 Methodology

We regard banks as assets having both a return and a risk component. By placing the banks in this risk-return framework, we analyze not only the profitability of the banks but also the risk inherent in profit maximization. When considering a portfolio of these assets and assuming that returns are not perfectly correlated, a supervisor holding this portfolio can diversify and enjoy a better risk-return trade-off. The diversification opportunities of the supervisor are constrained by the variety of banks available. Of course, banks themselves also individually diversify their loan portfolios, business lines or geographical markets served. However, the dimensions along which banks can diversify are correlated with their size, as larger banks can be assumed to have greater geographical reach and a different mix of activities compared with smaller banks. Moreover, if enough banks diversify along the same lines, the system as a whole becomes more susceptible to common shocks, even though individual banks themselves seem safer from a micro-prudential point of view. This observation is not new, being noted for instance by Rajan (2005), Wagner (2008), Acharya (2009), Ibragimov et al. (2011) and Allen et al. (2012), who also draw on Modern Portfolio Theory (Markowitz, 1952, 1959) and the insight that portfolio risk can be reduced by choosing assets that are not perfectly correlated with each other.

We contribute to this line of thinking by applying portfolio theory to the banking system in order to investigate the role of large banks in determining systemic (portfolio) risk. The supervisor in our case is able to change the portfolio's risk-return trade-off by altering the size of banks with respect to the system. Although this ability goes beyond the existing mechanisms in place (such as 'Cease and Desist' orders and other 'Prompt Corrective Actions'), we argue that this thought experiment can, at the very least, give us insights into the optimal design of a portfolio of banks. Applying portfolio theory to the banking system is not however all that straightforward, due to some of its strict assumptions. First, the return distribution is assumed to be fully characterized by the first two moments and disregards any tail dependence, even though financial market returns are found to be skewed and fat-tailed. To (partially) mitigate this issue, we use lower-frequency quarterly returns. Second, market participants are assumed to have no influence on prices and return structures of assets in their investment universe, regardless of the weight they are given. In reality, if a bank supervisor were to reduce the size of a bank, its risk-return trade-off would be bound to change as well, as would its correlation structure. Third, an investor is assumed to be able to purchase assets in parcels of any size, meaning that bank sizes could fluctuate heavily between the investor's decision moments. Under a more realistic scenario, the supervisor would be able to change the size of a bank in a limited way, e.g. by only a certain percentage of the bank's assets. Although we initially proceed under the strict setup, the last two assumptions are relaxed at a later stage.

In order to apply portfolio theory and build the regulator's portfolio, we first need to define the return and weight of the assets under consideration. Previous studies have relied mainly on market-based measures when assessing systemic risk (see e.g. De Jonghe, 2010; Adrian and Brunnermeier, 2011; Acharya, Engle and Richardson, 2012; Acharya, Pedersen, Philippon and Richardson, 2012; Bisias et al., 2012; Brownlees and Engle, 2012; Engle et al., 2014). Unlike these studies, we instead use the return-on-assets from book data for the returns of the banks. We do so for several reasons, the first being that the aggregated risk concerning the supervisor is not based on the returns and risks of the shareholders of banks, but rather on those of the (productive) assets that they hold. In the event that the regulator has to bail out a bank, saving or guaranteeing its liabilities will be equivalent to saving or guaranteeing its assets. Second, as shown by Allen and Carletti (2008), in financial crises market prices tend to reflect the amount of available liquidity instead of future earnings. Since these episodes are of particular interest to this analysis, market-based measures might not be appropriate as they could capture liquidity risk instead of systemic risk. Third, accounting data enable us to explore a more extensive sample since market data is only available for a small subset of banks. While listed banks do account for a large percentage of the total banking assets, small banks are potentially a source of (liquidity) contagion through the interbank market (see e.g. Furfine, 1999, 2003; van Lelyveld and Liedorp, 2006; Degryse and Nguyen, 2007). Finally, return-on-assets is a cleaner measure of the underlying profitability, as return-on-equity incorporates management choices with regards to leverage. While our baseline results are based on the book value of the return-on-assets, they are robust to using market-based measures.

The weight that the regulator holds in each bank is calculated as the bank-level total assets divided by the sum of all bank-level total assets available in the portfolio.

According to portfolio theory, the investor's return and risk are calculated as:

$$r_{p,t} = \mathbf{w}_t'\boldsymbol{\mu}_t \tag{6.1}$$

$$\sigma^2_{p,t} = \mathbf{w}_t'\boldsymbol{\Sigma}_t\mathbf{w}_t \tag{6.2}$$

where $\mathbf{w}_t$ is a column vector representing the weights of all banks in period t and $\boldsymbol{\mu_t}$ represents the expected return of the banks, usually defined as the average return of the previous quarters. Furthermore, $\boldsymbol{\Sigma}_t$ represents the expected covariance matrix of these returns and is often replaced by its sample equivalent. The average expected return of the portfolio is given by $r_{p,t}$, while the variance of this set of returns is given by $\sigma^2_{p,t}$ and represent the measure of portfolio (or in our case: systemic) risk at time t.

The supervisor is considered to be risk averse, and to prefer to hold the portfolio with the least amount of risk according to the objective function in Equation 6.2. Moreover, the supervisor is able to influence systemic risk by changing the weights in the portfolio, assuming that this does not impact the matrix $\boldsymbol{\Sigma}_t$.[1] Minimizing the objective function allows us to compare the differences in portfolio design between the initially realized and the hypothetical minimum variance portfolio.

To achieve the minimum variance portfolio (MVP), the supervisor solves for:

$$\begin{aligned} \underset{\mathbf{w}_t}{\operatorname{argmin}} \quad & \mathbf{w}_t' \boldsymbol{\Sigma}_t \mathbf{w}_t \\ \text{s.t.} \quad & \mathbf{w}_{i,t} \geq 0 \\ & \mathbf{1}' \mathbf{w}_t = 1 \\ & \mathbf{w}_t' \boldsymbol{\mu}_t \geq 0 \end{aligned} \tag{6.3}$$

with the addition of several further constraints. First, the supervisor cannot go short in a bank, i.e. no bank can have a negative weight. Second, the weights of banks have to add up to 1 as the existing assets are merely reshuffled, without any being created or destroyed. This is equivalent to assuming that the banks under consideration constitute the entire investment universe of the supervisor. Finally, a supervisor is also assumed to choose the weights such that the portfolio does not have negative returns.[2] It follows from these non-linear constraints that no analytical solution is possible, and we therefore rely on a numerical solution.

Using this approach, we investigate the following questions. First, we ask whether systemic risk can be reduced and if so, by how much. To do this, we compare the standard deviation of the initial portfolio with that of the MVP. Second, we investigate

[1] As the recent crisis has shown, it is not unusual for supervisors to intervene through liquidation, (hidden) bailouts or forcing banks to sell off assets to maintain a competitive environment.

[2] The inclusion of this constraint does impact our results as will be shown later.

Figure 6.2: The Federal Reserve Districts

whether the supervisor would have to sacrifice returns in order to achieve a lower risk, by comparing the average return of the initial portfolio with that of the MVP. Within the framework described in the introduction, we therefore ask if, in order to reach the MVP, the supervisor has to move along the efficient frontier or shift towards it. Third, we compare the dispersion of weights within each of the two portfolios, by looking at the share of the largest 5% of banks in the initial portfolio and in the MVP.[3] Finally, we ask whether the largest 5% of banks in the initial portfolio have retained their relative importance in the MVP by comparing their initial share with the weight they receive in the MVP.

6.2.2 Data

We perform our analysis on the U.S. banking system, which has several regulatory bodies at different levels. Depending on location, membership status and type, a bank can be regulated by the Federal Reserve System (FED), the Office of the Comptroller of the Currency (OCC) and the Federal Deposit Insurance Corporation (FDIC). While the main supervisory task at district level is carried out by the 12 Federal Reserve Banks, depicted in Figure 6.2, at the national level the main regulatory task is performed by the Federal Reserve's Board of Governors. As a consequence of this division, we consider the regulatory portfolio both at national level and at district level.

[3] An alternative measure of concentration would be the Herfindahl-Hirschman Index, although using the portfolio weights of the largest banks is more intuitive in this setting. The 5% concentration measure is preferred, since it allows for a better comparison in different-sized banking systems (see Alegria and Schaeck, 2009).

Bank data are obtained on a quarterly basis from the Call Reports for Income and Condition provided by the Federal Reserve System. For the national (FED) portfolio, we consider consolidated Bank Holding Companies (BHCs) as the assets in which the regulator can invest. Data for the BHCs are obtained from the FR Y-9C Forms, between 1986Q3 and 2012Q1. We select only Holding Companies and exclude Insurance/Securities brokers, Utilities and other Non-Depository institutions. At the Federal Reserve District (FRD) level, data on unconsolidated Commercial Banks are retrieved from the FFIEC 031/041 Forms between 1984Q1 and 2010Q4, excluding Savings/Cooperative/Industrial banks as well as Non Deposit Trust companies.

We use balance sheet data instead of financial market data, allowing us to consider all banks that are required to file reports and not only those that are listed on an exchange. Moreover, lower frequency returns are preferred to daily or even weekly returns, in order to comply with the assumption of normality of returns. We collect total assets (bhck/rcfd2170) and net income (bhck/riad4340), deflate both to 2005Q1 dollars using the Producer Price Index provided by the St. Louis Federal Reserve Bank, and filter out banks with return-on-assets exceeding +100% or -100%. This leaves us with 4,694 BHCs across 154,577 bank-year observations and 19,225 commercial banks over 1,132,425 bank-year observations.

Summary statistics for the BHCs and commercial banks are shown in Table 6.1. In any quarter in our sample, there are between 964 and 2,333 holding companies active in the United States. Due to inflation and a wave of consolidation, the total assets reporting threshold for BHCs was raised from \$150 million to \$500 million in 2006, causing a drop in the number of banks in the sample. Banks controlling less than \$500 million in total assets prior to 2006 are kept in the dataset, since, as will be shown at a later stage, their exit does not affect our results. Given that the distributions of assets and returns are highly unequal, we report percentiles instead of means and standard deviations. The median BHC controlled \$500 million in total assets and reported a net income of \$1.3 million. The table shows the skewness in the distribution of total assets, with the largest 5% of BHCs having total assets ranging between \$14.5 billion and \$2.1 trillion. While the net income of the median holding company is \$1.3 million, again there is a large disparity: the highest earning 5% of BHCs recorded profits ranging from \$44 million to \$6.4 trillion. At the other end of the spectrum, losses are equally

Table 6.1: Summary statistics

Panel A: Bank Holding Companies

	Min.	Percentiles 5th	25th	50th	75th	95th	Max.
Net Income (in $ million)	-15132.20	-0.72	0.64	1.30	3.09	44.0	6414.61
Total Assets (in $ million)	5.33	156.32	284.62	500.37	1184.04	14456.24	2115728.50
Return on Assets (in %)	-39.29	-0.14	0.18	0.29	0.39	0.58	82.81

	Mean	Std. Dev.	Min.	Max.
Number of Banks	1502	387	964	2333

Panel B: Commercial Banks

	Min.	Percentiles 5th	25th	50th	75th	95th	Max.
Net Income (in million $)	-11168.98	-0.20	0.08	0.24	0.58	2.94	4682.32
Total Assets (in million $)	0.14	15.91	39.48	79.39	171.95	896.35	1594746.30
Return on Assets (in %)	-78.23	-0.34	0.19	0.33	0.46	0.72	90.85

	Mean	Std. Dev.	Min.	Max.
Number of Banks	10115	2655	6477	14474

The table presents summary statistics for Net Income (bhck/riad4340), Total Assets (bhck/rcfd2170) and the Return-on-Assets of the banks in the analysis. Panel A displays these descriptives for Bank Holding Companies between 1986Q3 and 2012Q1, while Panel B summarizes them for Commercial Banks between 1984Q1 and 2010Q4. Due to the highly skewed distributions, we summarize the data according to their percentiles as well as the minimum and maximum values.

large, partly due to the recent financial crisis, with one holding company reporting a net loss of $15 trillion in the third quarter of 2008. Since the return-on-assets takes into account the size of the BHC, its values are less extreme compared with those of returns and assets separately, with the mean (0.257%), median (0.292%) and mode (0.325%) lying in close proximity.

The number of reporting commercial banks lies between 6,477 and 14,474 over the 12 districts. Regarding net income and total assets, commercials banks follow a similar pattern to that of Bank Holding Companies, although smaller on average. The total assets disparity is even larger than at the national level, with some banks dwarfing their competitors.

We proceed by placing every bank in its respective FRD portfolio, defining its weight as the total assets of the bank divided by the sum of total assets in the FRD. The BHCs are analyzed at the national level in a similar manner: the weights are calculated as the individual level of total assets divided by the sum of all total assets of the BHCs. As we have quarterly data over a period of 25 years at our disposal, we perform the analysis using a window of 8 consecutive quarters on which we calculate the expected return and sample covariance matrix, thereby taking into account time-varying correlation.[4] As a consequence, assets need to have posted data in each consecutive quarter of the window to be included in the analysis.

6.3 Results

In this section we present the results of our analysis. We first examine the risk-return trade-off between the FED portfolio and the MVP. In a second step, we look at the differences in portfolio allocation between the two systems before comparing their other features. We present the results of the analysis at BHC level graphically, referring the reader to the Appendix for the results on FRD level as they are quantitatively similar, and conclude this section with several robustness tests.

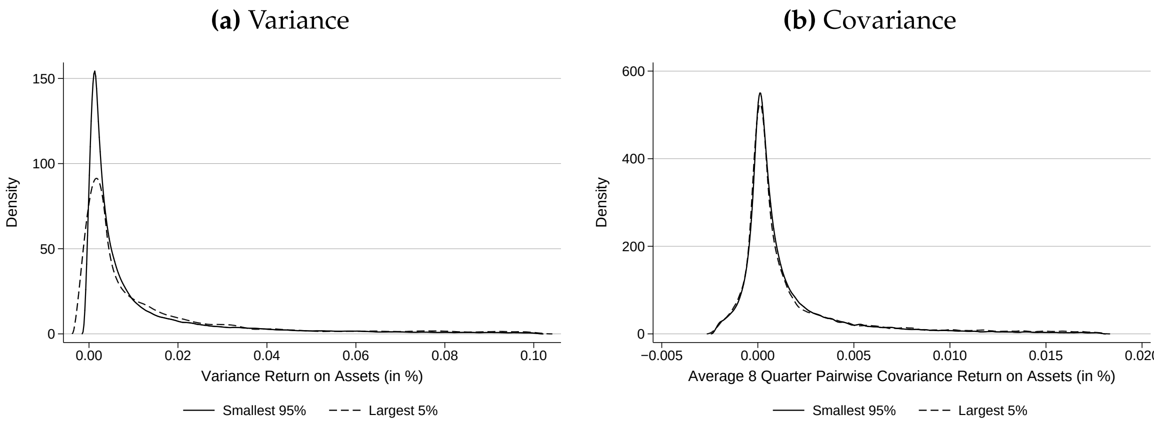
6.3.1 Are large banks more risky?

However, before reporting the results we first need to establish the similarities in the return structure between large and small banks. Should large banks have (co)variances different from those of small banks, the assumption that the covariance matrix is independent of size means we would impose an unrealistic structure when large banks are reduced in size, or small banks are made larger.[5] In Figure 6.3, we show the two dimensions of the covariance matrix by plotting the densities of the average 8 quarter variance and average 8 quarter pairwise covariance for the largest 5% and the smallest 95% of BHCs. From the Figure, it becomes clear that despite the differences in size, there is ample common support in the individual and common riskiness of bank returns as the distributions overlap almost entirely. These results hold for different time

[4] In Section 6.3.5 we show that our results are robust to a different window size.

[5] In this respect, our approach is similar to verifying whether the assumption of a ‘common support’ holds in propensity score matching (Heckman et al., 1998). If there is enough ‘common support’ between small and large banks, the assumption that the supervisor could change assets without having these actions leading to a different return structure is more realistic.

Figure 6.3: Are large banks more risky?



The figure shows density plots of the average 8 quarter variance of the return-on-assets and average pairwise covariances of the smallest 95% and the largest 5% of Bank Holding Companies between 1988Q2 and 2012Q1. Due to long tails for both large and small banks, the variance is truncated at its 90$^{\text{th}}$ percentile as it is strictly positive, while the covariance is truncated at its 5$^{\text{st}}$ and 95$^{\text{th}}$ percentile for graphical purposes.

samples, and for both BHCs and Commercial Banks.[6]

Of course, even if large and small banks share a common support in the covariance matrix, it does not mean that a bank that changes size will maintain its return structure. At a later stage, we therefore look at banks that have seen large increases or decreases in size, and analyze how this changed the elements of the covariance matrix. Using these average changes in turn allows us to alter the covariance matrix during the numerical optimization, leading to a more realistic portfolio allocation. However, since there are only a limited number of cases on which we can base this analysis, we first proceed by assuming that changing a bank's size does not influence the structure or level of its returns, and later revisit this assumption.

6.3.2 What role does inequality play in the risk/return trade-off?

It is quite straightforward to obtain the risk, return and weight distributions for the initial portfolios. By contrast, obtaining the respective MVPs is more cumbersome, as a minimum of 964 and a maximum of 2,333 BHCs are present during the sample period. The Chicago, Kansas City and Dallas portfolios typically contain well over

[6]Full results available upon request.

1,800 banks. The solution is computationally intensive, but is nonetheless obtained after a lengthy optimization process.

Figure 6.4 shows the results of the portfolio optimization based on BHC data: panels 6.4a and 6.4b display the risk-return trade-off for the FED portfolio and the MVP. They show that with different weights, portfolio risk is effectively eliminated in the MVP. Whereas the standard deviation of the FED portfolio spikes during the Savings-and-Loans and subprime mortgage crises, that of the MVP remains stable at around zero. One would expect a portfolio with lower (minimized) risk to have a lower return as well. However, this less risky system has positive returns throughout the sample period with values that closely track the actual returns. Therefore, we conclude that the initial portfolio does not lie on the efficient frontier, as risk is reduced while the level of returns has been maintained. Panel 6.4c shows that the lower risk is achieved in the MVP through a markedly lower concentration than in the FED portfolio. While in reality the weight of the largest banks lies between 65% and 90%, the concentration in the MVP is on average 13% and at most 52%. More interestingly, the largest banks in the FED portfolio see their cumulative weight reduced to at most 15% of total assets in the MVP.[7]

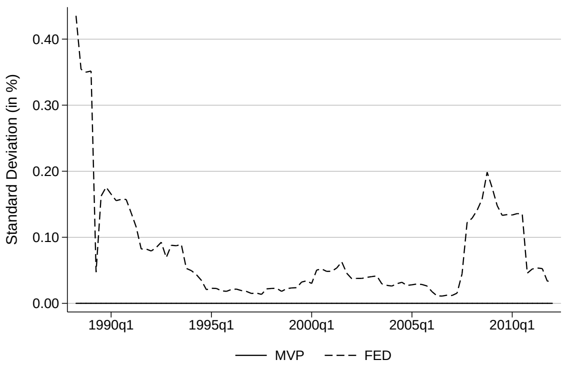
The same picture as for the BHC data emerges if we look at the separate FRD portfolios, reported in the Appendix. Table 6.A.1 summarizes the differences in risk and return between the FRD portfolios and their MVPs. Evidence of returns over the whole period is mixed: some FRDs outperform their MVPs, whereas others exhibit lower returns. One interesting fact is that in the boom period of 1994Q1 - 2006Q4, we find that all FRDs outperform their MVP counterparts in terms of returns. Regarding size disparity, Table 6.A.2 shows that the MVPs consistently have a much lower level of concentration compared to their FRD portfolios, the difference ranging on average between 44% and 78% throughout the sample period.

Two further remarks are in order. First, the portfolio standard deviation in Figure 6.4 seems to be higher during the S&L crisis than in the subprime crisis. This result can be explained by the fact that a standard deviation, unlike a correlation coefficient, is not a dimensionless number and can only be interpreted as a function of its mean.

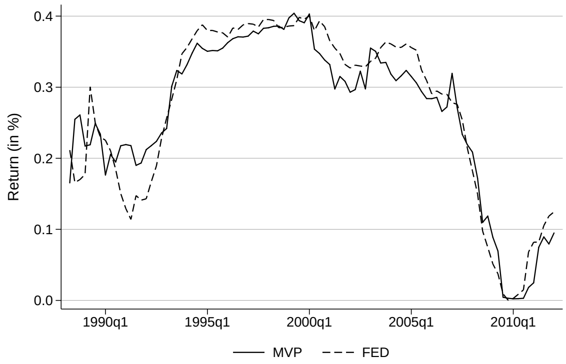
[7] Although not reported in the Figure, the Gini coefficient exhibits the same pattern, as the average coefficient of the FED portfolio is 0.9 compared to 0.2 for the MVP.

Figure 6.4: What role does inequality play in the risk/return trade-off?

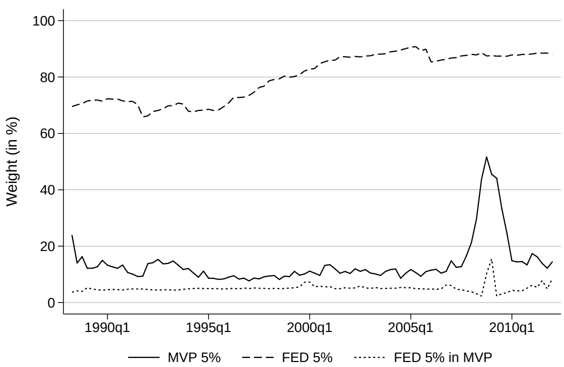
(a) How much is systemic risk reduced in the MVP?



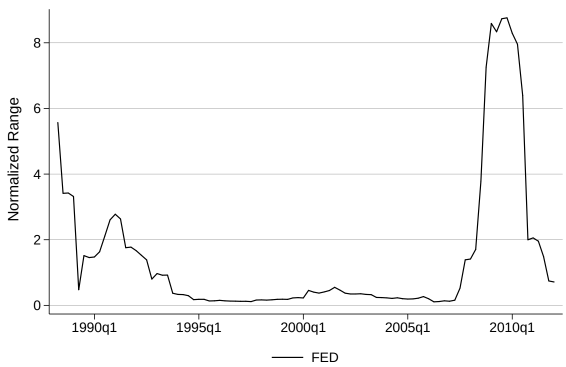
(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



(d) How does the systemic risk measure evolve over time?



The figure shows the comparison between the FED portfolio and the hypothetical MVP: panel 6.4a and 6.4b display the difference in the risk and return of each portfolio. Panel 6.4c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP. Panel 6.4d shows the demeaned risk measure from Panel 6.4a to allow for a comparison over time. Instead of using the standard deviation, which is a reflection of the mean, we show the normalized range defined as (max-min)/median and is independent of the mean return.

Given that the returns of the FED portfolio and the MVP are of similar size throughout the sample, we can compare their standard deviation in each quarter, but not between quarters. To allow for comparison over time, we de-mean the risk measure, although this leads to a loss in direct interpretation.[8] Figure 6.4d displays the de-meaned risk measure, and portfolio risk now shows a higher peak in the subprime crisis than in the S&L crisis.

The second issue we want to address is the MVP's high concentration during 2008 and 2009. This spike can be explained by the fact that up to 87% of BHCs reported lower average returns than in the previous quarter and 40% of all BHCs recorded losses. Given the number of banks involved, it is possible that concentration rose because of this increase in correlation between average returns. However, it is also possible that the MVP weights are chosen to avoid violating the no-loss constraint. To test the latter possibility, we ran the analysis excluding the no-loss constraint but still find the same spike, indicating that, indeed, higher weights are given to banks that share a lower correlation. Since these are few in number, they therefore have to receive a higher weight in order to minimize portfolio risk.

The results here suggest that inequality and concentration play an important role in the risk-return trade-off with which a regulator is faced. In this simple exercise, reducing inequality drives down risk without significantly affecting returns at both FED and FRD level. These findings indicate that regardless of the regulatory level, supervisors need to be concerned when looking at the optimal design of their portfolio not only with a bank's individual size but also its size relative to the system. Moreover, the rebalancing of weights does not appear to be random. We find that in order to obtain a less risky portfolio, a supervisor has to reduce holdings of the currently largest banks and create a more equal system. In reality, the largest banks have had a much higher share in the portfolio compared with that in the MVP, and even increased their weight from 65% to 90% during the sample period. Moreover, the financial industry as a whole has also grown in relation to GDP, to the extent that Carvalho and Gabaix (2013) attribute the recent rise in macroeconomic volatility mainly to this growth in combination with idiosyncratic shocks to the largest banks. Indeed, the share of the

[8] We do so by calculating the normalized range of returns. Since the portfolio standard deviation is the standard deviation of the weighted average returns during the last 8 quarters, we define the normalized range as the (maximum-minimum)/median of this set.

largest banks in the current portfolio relative to GDP has increased from 35% at the beginning to 75% at the end of the sample. Given the evidence presented here, combined with the finding of Carvalho and Gabaix (2013), large banks do seem to play an increasingly important role not only in the banking system, but also in the broader economy and its volatility.

However, one important caveat needs to be acknowledged. In the methodology section, we assumed that the banks under consideration represent the entire investment universe for the supervisor, mimicking the current regulatory set-up. Therefore, we exclude financial institutions that fall outside of the regulator's jurisdiction, such as for instance investment banks and money market funds that played a big role in the propagation of risk during the 2008 crisis. Similarly, while U.S. subsidiaries of international banks have to file financial statements and are therefore included, their international parents are not considered. While on a global scale their activities can impact the profitability and risk of the banks under consideration, we are only able to observe their indirect effect on the U.S. system. As a consequence, if the supervisor would set his weights according to the minimum variance portfolio, he might force banks to move their activities abroad or into the shadow banking realm where they are unsupervised. Given the data at hand and our assumption that assets are redistributed and not moved outside of the portfolio, we can not consider this outcome. The results presented here and in the following sections therefore correspond only to a *partial* equilibrium.

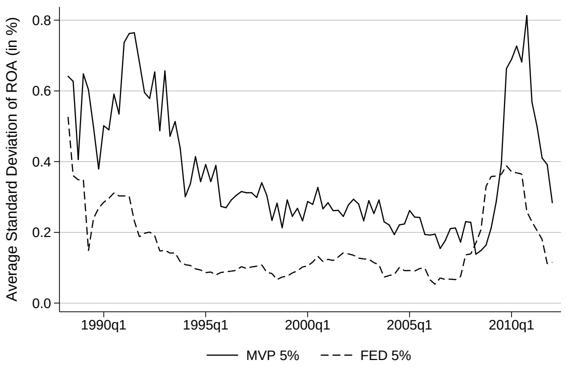
6.3.3 What are the other features of the Minimum Variance Portfolio?

Given that we have seen that lower portfolio risk is achievable in a less concentrated banking system, this raises the question of what causes the largest banks in the FED portfolio to have such consistently low weights in the MVP, and how they differ from the largest banks in the MVP. The two components that determine the weighting decision are on the one hand individual bank risk, as measured by the standard deviation of the returns, and on the other hand the correlation between these returns.[9] Figure 6.5a shows the average standard deviation of return-on-assets for the largest 5% of

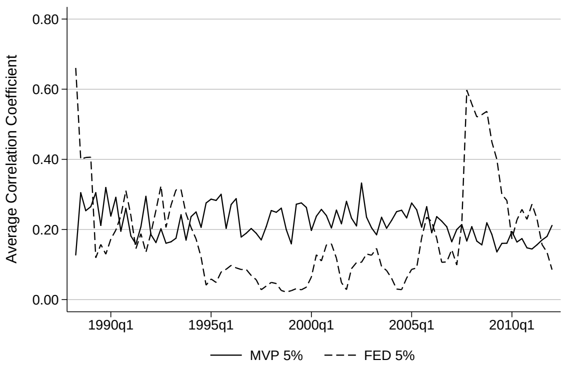
[9] In this section, we show the correlation of returns instead of the covariance as it is easier to interpret.

Figure 6.5: What are the other features of the Minimum Variance Portfolio?

(a) Is there a large share for low-risk banks?



(b) Is there on average less correlation among banks?



The figure shows the average bank level standard deviation of the return-on-assets, as well as the average pairwise correlation between these returns for the largest 5% of banks in both MVP and FED portfolio.

banks in the FED portfolio and MVP, while Figure 6.5b displays the average pairwise correlation among the largest 5% of banks in their respective portfolios. We see that the standard deviations of the returns of these largest banks are on average twice as high in the MVP, yet the average correlation coefficient is much more stable compared to the FED portfolio. In the two crisis periods, the FED portfolio pairwise correlation spikes to average values of 0.6, almost three times larger than the MVP. Nonetheless, even when the individual risk and correlation are lowest in the FED portfolio, its risk is still higher than that of the MVP.

These observations add to the evidence that in this context, weight plays a significant role in determining the level of risk, as it magnifies the effect of increased correlation and individual riskiness. When both components have low values, systemic risk is low, even in a highly concentrated market. However, when they increase in crisis periods, systemic risk increases dramatically if size inequality is high. As already shown in Gabaix (2011), individual shocks to firms have the potential to lead to aggregate volatility when the size distribution of an economy is heavy tailed, something that also holds for the banking system (see e.g. Janicki and Prescott, 2006; Blank et al., 2009). In terms of portfolio theory, the variances will dominate the covariances in crisis periods due to the large disparity in weight. Since the movements in correlation can be extremely volatile and difficult to control or even predict, the best tool for keeping systemic risk low in this context is to limit concentration.

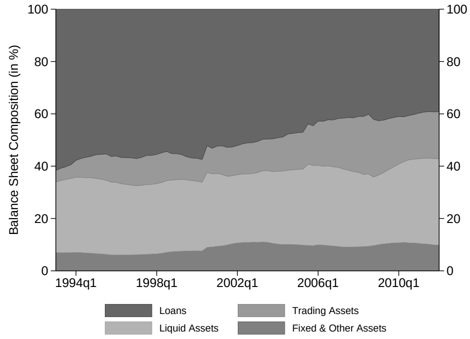
We note, however, that these results do not necessarily imply a cap on the size of banks. On the one hand, a system with only small banks is subject to the Too-Many-To-Fail problem since they could herd, thereby acting as one large entity (Acharya and Yorulmazer, 2007; Claeys and Schoors, 2007; Brown and Dinç, 2011). In this setting, herding would be picked up via an increase in correlation of returns, posing a systemic threat despite a lack in concentration. Our results, however, indicate that the correlations of small banks are relatively stable over the sample period, and would therefore not pose a systemic threat. On the other hand, the system can also be diversified by limiting activities that banks can undertake and/or markets it can serve, provided they operate in their own (uncorrelated) niche. This point was also touched on by Loutskina and Strahan (2011), who found that increased geographic diversification went hand-in-hand with a decline in loan monitoring by lenders prior to the financial crisis.

To determine the characteristics of banks which have been heavily reweighted, we construct a crude industry level balance sheet for both the FED portfolio and the MVP. We use the weights allocated to each bank to construct this weighted average balance sheet, which is shown in Figure 6.6. The allocation of assets in the FED portfolio shows the increasing importance of trading assets at the expense of loans, whereas this trend is less evident in banks favored in the MVP. While the FED balance sheet has less than 40% of assets invested in loans at the end of the sample, that of the MVP remains close to 50%. On the liabilities side, the FED portfolio is more reliant on non-deposit funding than the MVP balance sheet. We observe that at the end of the sample, the FED portfolio uses about 10% more of these non-deposit liabilities than the MVP, although this difference was much more apparent before the recent crisis.

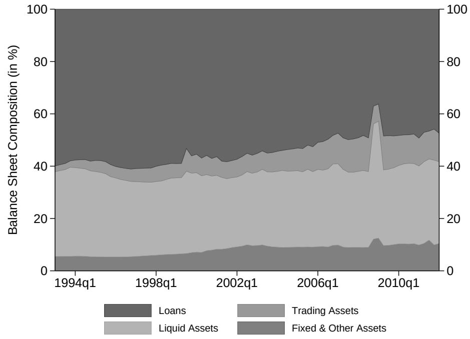
As a reflection of the industry asset composition, the non-interest income/total income ratio for each portfolio is shown in Figure 6.6e. We observe that with exception of the crisis years, there has been a significant increase in reliance on non-interest income in the FED portfolio. On the other hand, banks favored in the MVP have a more constant share of non-interest income throughout the sample period. Notwithstanding the financial crisis, the gap between the portfolios has been steadily increasing since 1990.

Figure 6.6: What happens to the intermediary role of banks in a safer banking system?

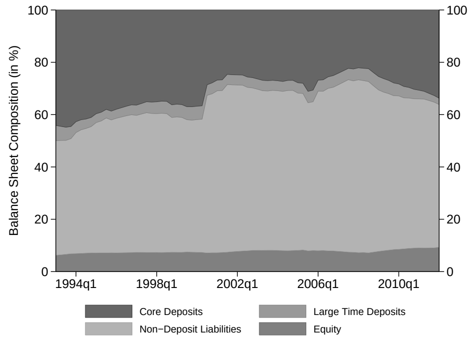
(a) FED - Assets



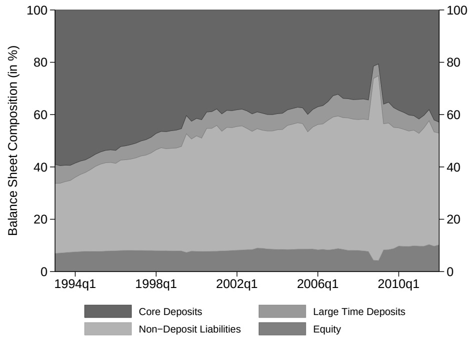
(b) MVP - Assets



(c) FED - Liabilities



(d) MVP - Liabilities



(e) Is there a difference in income characteristics?

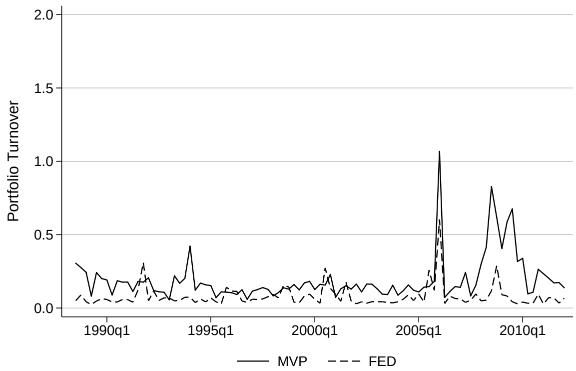
The figure compares the intermediary role of the FED portfolio and MVP, by showing weighted average industry level balance sheets in panels 6.6a-6.6d and the different weighted non-interest income/total income ratios in panel 6.6e. To construct the balance sheet we use the following data series: Loans (BHCK2122); Trading Assets up to 1994 (BHCK2146); Trading Assets after 1994 (BHCK3545); Liquid Assets up to 1994 (BHCK0081 + BHCK0395 + BHCK0397 + BHCK3365 + BHCK0390); Liquid Assets after 1994 (BHCK0081 + BHCK0395 + BHCK0397 + BHCK3365 + BHCK1754 + BHCK1773); Fixed & Other Assets (BHCK2145 + BHCK3163 + BHCK2160); Core Deposits (BHCB3187 + BHCB2389 + BHCB6648 + BHCB2210 + BHOD3187 + BHOD2389 + BHOD6648); Large Time Deposits (BHCB2604 + BHOD2604); Non-Deposit Liabilities (BHCK2948 - Core Deposits - Large Time Deposits); Equity (BHCK2948 + BHCK3210). Non-interest income (BHCK4079); Total income (BHCK4107).

Given these results, we conclude that the fictitious banking industry in the MVP is characterized by retail banking, as higher weights are given to banks that are mainly funded by deposits, make loans, and therefore rely less on non-interest income. Since banking activities are not a direct input in the minimization of the portfolio risk, it is their influence on the behavior of the returns which drives these findings. Indeed, this is in agreement with a growing literature emphasizing the role of income diversification in financial instability. For instance, Stiroh (2004) and Stiroh and Rumble (2006) find that non-interest income reduces aggregate profits while increasing risk. More recent evidence by De Jonghe (2010) shows that systemic risk is exacerbated by banks diversifying into activities other than lending, due to increasing correlations between income streams. This finding was also corroborated by Adrian and Brunnermeier (2011), Brunnermeier et al. (2012), DeYoung and Roland (2001) and DeYoung and Torna (2013). Huang and Ratnovski (2011), meanwhile, argue that wholesale lenders have lower incentives for costly monitoring, leading to large (and inefficient) fluctuations of loans on negative public signals, a problem not encountered in relationship banks. Finally, Boot and Ratnovski (2012) find that although there are initial benefits for banks from starting trading activities, beyond a critical point inefficiencies dominate and trading becomes increasingly risky. On the funding side, Fahlenbrach et al. (2012) emphasize that banks with increasing balance sheets through the use of short term non-deposit liabilities performed poorly during the last crises.

6.3.4 How easily is the Minimum Variance Portfolio obtained?

In our baseline scenario, the supervisor is able to switch assets rapidly from one bank to another on a quarterly basis to obtain the MVP. Although reweighting also occurs naturally in the FED portfolio via mergers and acquisitions, bank entry and exit or bailouts, the MVP would not be a realistic approximation if reweighting was much higher than in reality. In order to assess how stable the MVP is over time compared with the FED portfolio, we therefore calculate both of their turnovers. Turnover is defined as the sum of absolute weight changes in the portfolio between period $t-1$ and t, taking values ranging from zero (no change) to two (where all assets that were held are sold, and all assets that were not held bought). Figure 6.7 plots the turnover for both portfolios. We observe that the MVP turnover is on average 3 times as high

Figure 6.7: How much more intervention would be required for the MVP?



The figure shows the difference in reweighting between the FED and the MVP by plotting the turnover of each portfolio.

as that of the FED portfolio. The spike in turnover in 2006Q1 is due to changes in the reporting threshold, as the banks that reported in 2005Q4 were considered to be sold in 2006Q1 and proceeds reinvested in the other banks. The MVP can only achieve low risk through a higher level of reweighting, especially in the crisis period. For the district portfolios, the average MVP turnover is around 3 times higher than the actual portfolio, ranging from 2 times for the least concentrated to 6 times for the most concentrated districts. In Section 6.4, we therefore explore alternative MVPs where the reweighting is restricted, in order to achieve a more realistic turnover.

6.3.5 Robustness

To find out how robust our results are, we test several of our assumptions. The results of these tests are summarized in Table 6.3, where we evaluate how well risk was reduced while limiting concentration. To this end, we define the ratio $(\sigma_{\text{FED}} - \sigma_{\text{MVP}})/\sigma_{\text{FED}}$, which measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements were possible. A test is regarded as successful when this ratio averages 0.9 or higher throughout the sample period, and when the level of concentration of the largest 5% of banks in the MVP is on average below 50%.

Covariance matrix – 1 We begin with the assumption that the covariance matrix is in-

Table 6.2: Changes in the return and its covariance matrix

		ROA		Variance ROA		Covariance ROA	
		Pre-Jump	Post-Jump	Pre-Jump	Post-Jump	Pre-Jump	Post-Jump
Positive Jump	Value	0.341	0.259	0.023	0.111	0.002	0.005
	Average Difference	-0.082		0.088		0.004	
	Percentage Difference	-23.97%		379.70%		233.70%	
	P-value T-test	0.000		0.007		0.010	
Negative Jump	Value	-0.142	0.213	0.281	3.249	0.007	-0.015
	Average Difference	0.355		2.967		-0.022	
	Percentage Difference	249.52%		1054.64%		-324.52%	
	P-value T-test	0.110		0.359		0.293	

The table shows the average differences in ROA, its variance and its average pairwise covariance pre- and post-jump for both negative jumps and positive jumps using the procedure described in the text and the Appendix.

dependent of the size of banks. In Section 6.3.1 we showed that large and small banks share a common support in the variance and average pairwise covariance. This however, does not imply that a bank which changes in size will maintain the same level of returns or the same structure with regard to other banks. If we knew how the return structure changes due to a change in size, we would be able to adjust the covariance matrix in each iteration of the optimization. To this end, we have identified 15 cases in which BHCs experience a negative jump in bank size, and 287 where they experience a positive jump in bank size.[10] A jump is defined as an increase/decrease of bank assets of 25% or greater from one quarter to the next, provided that the preceding and following 8 quarters did not show jumps larger than 10% in each of the quarters, nor a cumulative change in the preceding and following 8 quarters of 25%. These last two conditions are imposed to make sure that bank size before and after the jump was relatively stable and that the change in the elements of the covariance matrix can be chiefly attributed to the one-time jump. The banks receiving a negative shock lost 40% of their total assets on average, while banks receiving a positive shock gained 60%. The average changes in the return and covariance matrices are displayed in Table 6.2. T-tests show that banks experiencing a positive jump in assets have a statistically significant lower return-on-assets, which is likely due to the construction of the variable, and a higher average variance and covariance of these returns. On the other hand, banks experiencing a negative jump do not see changes in their average variance and covariance, and only see a marginally significant higher return-on-assets.

Using the statistically significant changes for positive and negative jumps, we incorporate the effect in the sample covariance matrix and the return matrix, such that these matrices change dynamically with the weights of the banks.[11] We perform two additional robustness tests. In the first test, the vector of expected returns is adjusted according to the changes in weight, and a new sample covariance matrix is estimated with which portfolio risk will be minimized. In the second test, we adjust both the return and covariance matrix based on the changes in weight. Changes in the return and covariance matrices are interpolated if the proposed change in portfolio weight lies between -40% and +60%. As there was no data on changes in assets larger than these bounds, we use any proposed changes beyond them as if they were -40% or +60%, i.e. while a bank can receive an increase in weight higher than +60%, its variance and covariance terms are adjusted as if the weight has only been increased by +60%.[12] In both cases, however, this adds more complexity to the optimization and indeed we find that there are cases in which no improvement in portfolio risk is found. Fortunately an improved solution is still possible in most of the sample period, as can be seen in Figure 6.8. Interestingly, the findings in the baseline specification seem to be robust and are not influenced by our original assumption. The concentration in the MVP remains much lower than that of the FED portfolio, while simultaneously maintaining a smaller portfolio risk and returns of a similar level.

Covariance matrix – 2 Another issue in our baseline setup is that optimization can be quite unstable when using the sample covariance matrix: small changes in the return structure can lead to large differences in the outcome of the portfolio choice. To counter these unstable solutions, Brodie et al. (2009) use regularization of the optimal mean-variance portfolio by including a penalty in the objective function. They show that while introducing this penalty can lead to a sparse choice of weights, opting for a high penalty is equivalent to a constraint that does not allow shorting. Since our supervisor is already assumed not to be able to go short in any banks, we therefore consider the

[10] Full details on this identification can be found in the Appendix.

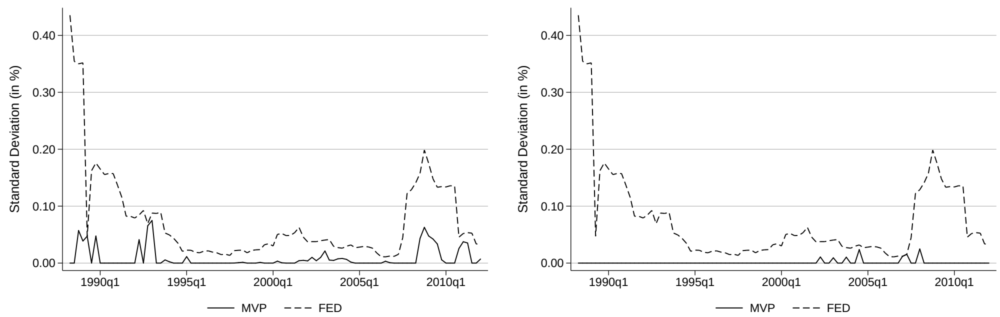
[11] As the change in return-on-assets for the banks experiencing a negative jump is marginally significant at 11%, we also regard this change as significant.

[12] A scenario under which reweighting of each bank was limited to -40% and +60% yielded no improvements over the actual portfolio. Results are available upon request.

Figure 6.8: Can portfolio risk be minimized when taking into account a changing covariance matrix due to weight changes?

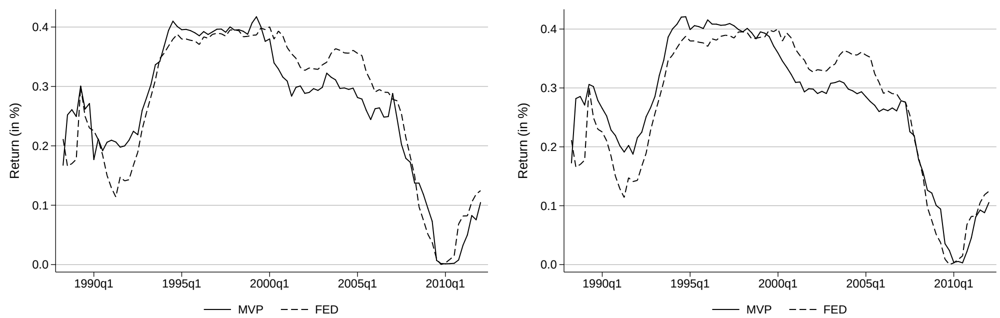
How much is systemic risk reduced in the MVP?

(a) Dynamic Return Matrix **(b)** Dynamic Covariance Matrix



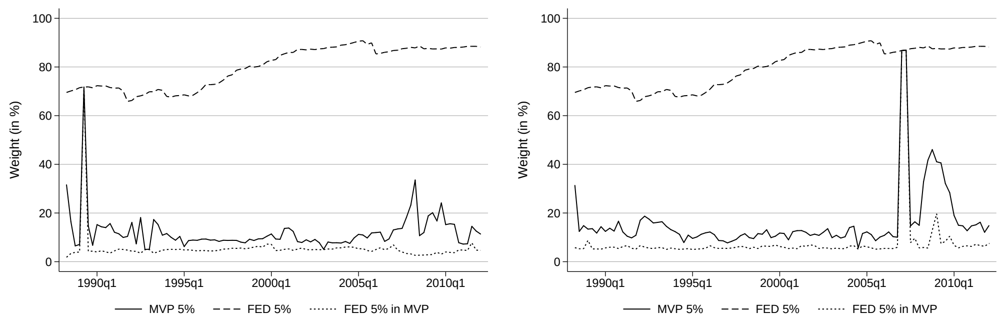
Do returns have to be sacrificed to achieve this lower risk?

(c) Dynamic Return Matrix **(d)** Dynamic Covariance Matrix



How unequal is the MVP compared to the FED portfolio?

(e) Dynamic Return Matrix **(f)** Dynamic Covariance Matrix



The figure presents two robustness tests, in which the return and/or covariance matrices are dynamically updated. In the first robustness test (Dynamic Return Matrix), we adjust only the return matrix and then estimate a new sample covariance matrix. In the second test (Dynamic Covariance Matrix), we adjust both the return matrix and the covariance matrix with which the portfolio risk will be minimized. More information on how we perform these tests can be found in the Appendix. Panel 6.8a and 6.8b display the difference in the risk of each portfolio, while panel 6.8c and 6.8d display the difference in return. Panel 6.8e and 6.8f show how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

optimization problem stabilized. However, as Jobson and Korkie (1980) have pointed out, a considerable amount of noise is introduced in a sample covariance matrix in a small T and large N setting like ours. In response, Ledoit and Wolf (2003, 2004) proposed a shrinkage based estimator of the covariance matrix to reduce this noise. We use their proposed estimator of the covariance matrix instead of its sample equivalent and re-run the analysis.[13] The full results are reported in the Appendix and are similar to the baseline specification. We find that the standard deviation of the MVP is now higher compared with the baseline specification, although still lower than that of the FED portfolio, while returns are at a similar level. The levels of concentration in the MVP are basically unchanged, as the largest banks on average still have a weight of 10%, while the largest banks in the FED portfolio have a weight in the MVP of 5%.[14]

Starting values Third, we explore the optimization starting values and choice of the length of the moving window. Given the fact that we are dealing with many banks, the minimization of the portfolio risk is likely to be a complex, highly nonlinear problem comprising multiple minima/solutions. The starting values, which are selected as the weights in the original portfolio, can have a substantial impact on whether a global or local minimum is found and in what direction the distribution of weights will move. To account for this possible bias, we run two robustness tests. In the first, we choose starting values based on an equally weighted portfolio. In the second, we run, for the BHC data only, 100 repetitions per quarter using randomized starting values.[15] Both tests show that the results are almost identical to the baseline specification, and we therefore refer the reader to the Appendix for the full results.

Length of rolling window Finally, we explore alternative lengths of the rolling window. So far we have taken an 8 quarter time frame to estimate the sample covariance matrix. However, it could be argued that using more data to estimate it would be less noisy and less prone to outliers. Taking this into account, we rerun the analyses

[13]The code for estimating the covariance matrix is obtained from: `http://www.ledoit.net/honey_abstract.htm`

[14]As portfolio risk is not minimized by at least 90%, however, we do not consider this test to be successful in Table 6.3.

[15]The random starting values are drawn from a half-normal distribution and then divided by its sum, such that they add up to 1.

Table 6.3: Robustness tests

	Issue	Robustness test	Risk minimized without large banks? BHC	COMM
(1)	Independence of covariance matrix	Dynamic return matrix	✓	n.a.
(2)	Independence of covariance matrix	Dynamic covariance matrix	✓	n.a.
(3)	Noise in sample covariance matrix	Shrinkage estimator	✗	0/12
(4)	Multiple minima	Equal starting weights	✓	12/12
(5)	Multiple minima	Randomized starting weights	✓	n.a.
(6)	Length of rolling window	Analysis on 16 quarters	✓	12/12

The table presents a summary of the three robustness tests that were performed, indicating whether portfolio risk was successfully minimized while keeping the levels of concentration of the largest 5% of banks low. To see if a test is successful we define the ratio $(\sigma_{\text{FED}} - \sigma_{\text{MVP}})/\sigma_{\text{FED}}$, which measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements were possible. A test is regarded as successful when this ratio averages 0.9 or higher throughout the sample period, and when the level of concentration of the largest 5% of banks in the MVP is below 50%. The second, fifth and sixth tests are only performed on BHC data due to their computationally intensive nature.

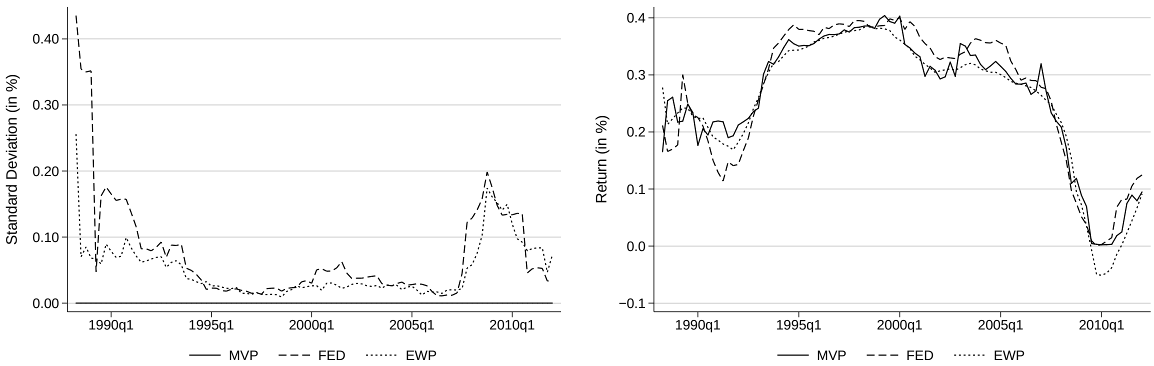
using a 16 quarter window and report the full results in the Appendix. For both the BHCs and commercial banks, results follow similar patterns to those using 8 quarter windows: the largest banks are still shown to be consistently overweighted compared with their MVP counterparts, where lower risk is achieved while keeping returns at a comparable level. Table 6.3 summarizes the results of the robustness tests in this section.

6.4 Extensions and Policy implications

As we have seen in Section 6.3.4, the amount of turnover needed to lower systemic risk is three times higher in the MVP. In this section, we therefore explore some more realistic scenarios, and discuss implications for policy resulting from the analysis. We first look into several weighting alternatives. Besides analyzing these other weighting methods, we repeat our analysis on a smaller and more realistic sample of banks. Finally, we discuss whether optimization at district level also results in a lower countrywide systemic risk. Similar to Section 6.3.5, we summarize all results in Table 6.4, where we again define a test successful if the ratio $(\sigma_{\text{FED}} - \sigma_{\text{MVP}})/\sigma_{\text{FED}}$ averages 0.9 or higher throughout the sample period.

Figure 6.9: What is the risk/return profile of an equally weighted portfolio?

(a) Can we reduce systemic risk in an equally weighted portfolio? **(b)** Is an equally weighted system more profitable?



The figure presents results of the risk and return in an equally weighted portfolio, compared to the MVP and FED portfolio.

6.4.1 Can portfolio risk be minimized while limiting portfolio turnover?

Equally weighted portfolio Given that the MVP seems to favor a more equal distribution of assets, a natural course of action would be to analyze an equally weighted portfolio. As noted by DeMiguel et al. (2009), equally weighted portfolios still outperform many optimizing portfolio choice models and have a very low turnover. The turnover in our setup would indeed be lower than that of the FED portfolio, albeit not zero as bank entry and exit would still take place. Figure 6.9 shows the risk-return trade-off that the equally weighted portfolio (EWP) would have in comparison with the other two. In terms of returns, the EWP performs similarly to the MVP and FED portfolio, except for the last crisis period in which it records losses. Regarding risk, the EWP has levels similar to that of the FED portfolio, albeit marginally lower. All in all, this suggests that there is an optimal level of concentration, as neither a highly concentrated nor an equally weighted portfolio are able to significantly reduce systemic risk in the same way as the MVP.

Limited reweighting – 1 Since a high turnover is costly for the supervisor and therefore not very desirable, we consider several alternatives involving limited reweighting that could reduce turnover. We do this by setting lower and upper boundaries to the weights banks can take, conditional on their true weights. First, we allow banks to grow/shrink by 10% and 20% of their initial weight. Second, we construct a measure of asset growth in the previous quarter and allow changes equal to either the mean or standard deviation of this growth measure. Whereas the first constraint is static in nature, the second allows for business cycle effects to determine how much reweighting can take place. To ensure the no-loss constraint does not influence the results, we run the limited reweighting scenarios with and without this requirement. However, regardless of the specification, the risk in the MVP is practically unchanged compared with the FED portfolio in all time periods.[16]

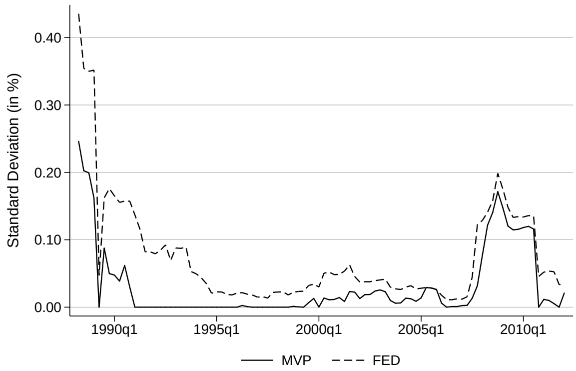
Limited reweighting – 2 Another way of reducing turnover would be to keep the largest 5% of banks at their current cumulative size, allowing unlimited reweighting of the remaining banks while still adhering to the no-shorting and no-loss constraint. The results of this exercise are shown in Figure 6.10, where we see that for most of the sample period it was possible to decrease systemic risk significantly while maintaining large banks. However, during the S&L and subprime crises periods, this MVP variant has a risk which is barely below that of the initial portfolio. This finding can be explained by the high correlation of returns for the largest banks, as seen in Figure 6.5b. During sudden increases in correlation between these largest banks, the high concentration of assets in a few banks will inevitably affect risk. In terms of optimal portfolio design, it seems to point to a trade-off between concentration and correlation: if the supervisor wants to keep the large banks at their current size, it would be necessary to ensure that correlation between them remains relatively low to avoid the Too-Many-To-Fail problem. This could, for instance, be achieved by limiting the geographical markets in which a bank can be active or the activities it can engage in, as used to be the case prior to e.g. the Riegle-Neal and the Gramm-Leach-Bliley acts.

Repeating this analysis on the FRD portfolios shows that there are 3 districts which are able to reduce systemic risk to a minimum in each of the three sub-samples.[17] Interest-

[16] Since the standard deviation of asset growth was extraordinarily large in 1997Q4, this allowed the optimization to apply larger changes to the banks and therefore managed to reduce risk. A dynamic approach was also considered for the scenarios where banks are allowed to grow/shrink by 10% and 20% of their weights in the MVP in time $t-1$; however, portfolio risk was not significantly reduced. Full results can be found in the Appendix.

[17] Full results are shown in the Appendix.

Figure 6.10: How much can we lower systemic risk when we keep the largest banks at their actual size?

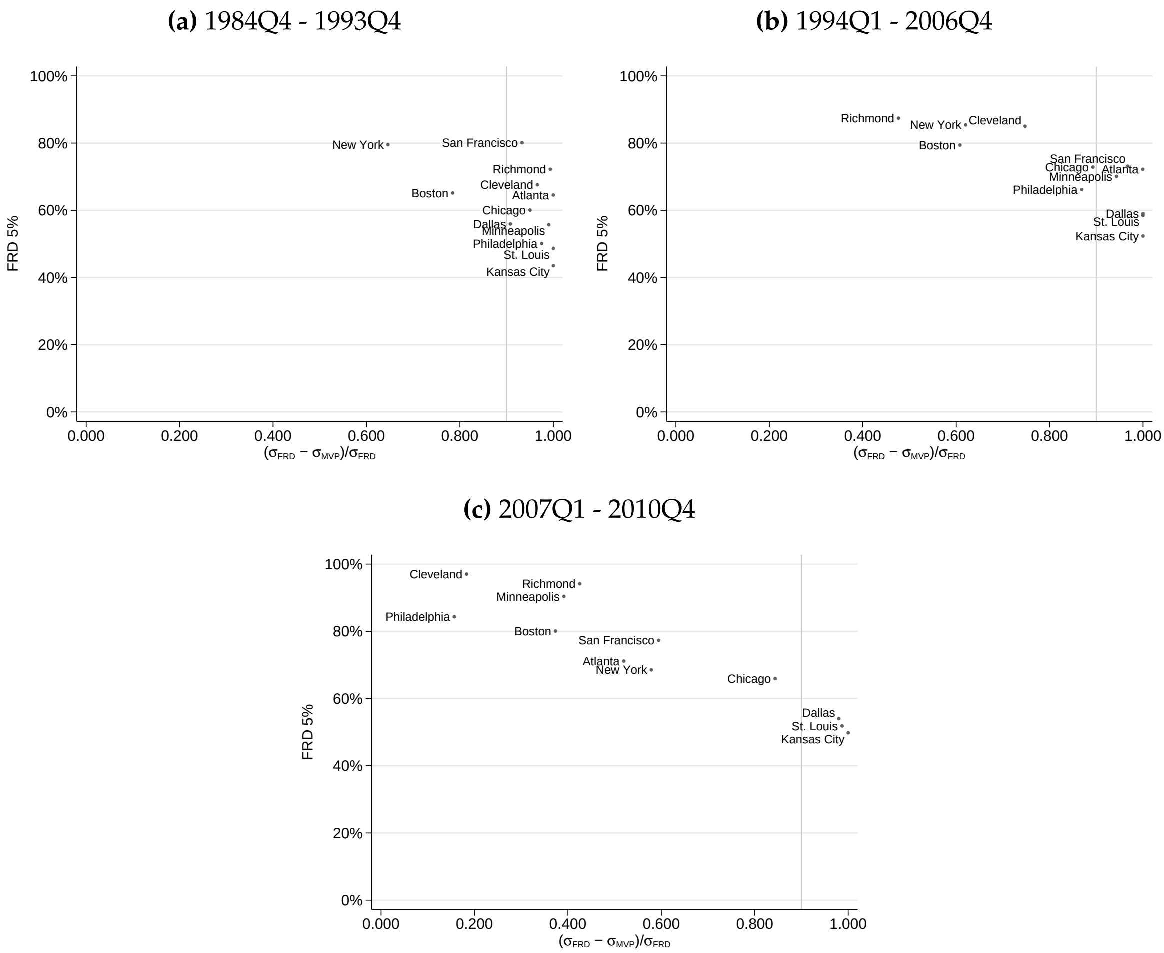


The figure displays how well the portfolio risk can be minimized while keeping the largest 5% of banks at their actual size and reweighting the remaining 95%.

ingly, a common feature is that they have the lowest levels of concentration among all districts. This relationship is shown in Figure 6.11, where we plot the average concentration ratios against the extent to which they are able to reduce systemic risk in three time periods. The horizontal axis represents the amount by which they reduce risk and is again constructed such that 1 stands for a reduction of risk effectively to zero, and 0 indicates that no improvements in risk are possible. The Figure clearly shows the trend in consolidation, with most districts becoming more concentrated over time. As they become more concentrated, they find themselves less able to achieve low risk while maintaining their largest banks.

Limited reweighting – 3 Since the weight of the 5% largest banks in the three successful districts never exceeds 60%, this leads us to a final test using limited reweighting. Is it possible to reduce portfolio risk while maintaining the cumulative weight of the largest 5% of banks between 50% and 60%? The results for the FED portfolio are plotted in Figure 6.12. We observe that under this limited reweighting scheme, it is possible to effectively eliminate risk while maintaining similar returns. Concerning the concentration in the portfolio, the share of the largest banks always hits the lower bound of 50%. Consequently, the level of concentration in the MVP also lies close to 50% except for 2008 and 2009, with similar results for the analysis at commercial bank level as shown in the Appendix. As in the previous scenarios, the high concentration in the MVP in 2008 and 2009 is likely due to increasing correlation of returns. To mini-

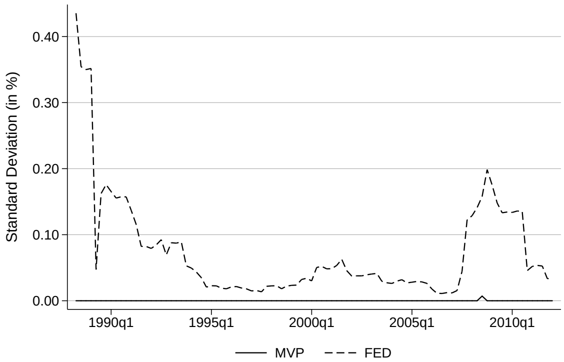
Figure 6.11: Can inequality explain why some FRDs can reduce systemic risk while keeping the largest banks at their actual size?



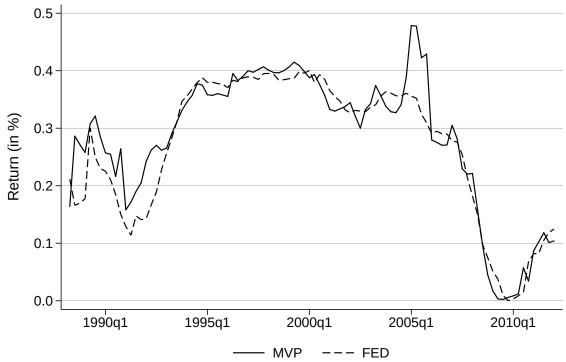
The figure shows the relationship over time between the average weight of the largest 5% of banks and the ability of the FRDs to reduce portfolio risk when the largest banks are allowed to keep their initial weight. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements were possible. The threshold for successfully minimizing portfolio risk is set at 0.9. The weights of the largest 5% of banks are averaged over each time period.

Figure 6.12: By how much is the risk minimized when the largest banks hold between 50% and 60% of assets?

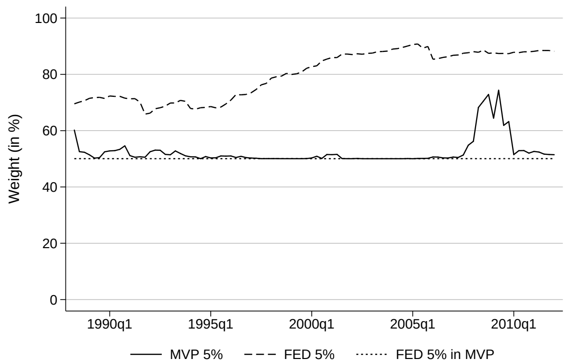
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How large are the largest banks in the MVP?



The figure shows the risk and return characteristics of an MVP where the largest 5% of banks in the FED portfolio are kept between 50% and 60% of assets, and how the weights in this MVP are distributed.

mize portfolio risk, a small number of banks need to receive a higher weight, such that their concentration approaches that of the FED portfolio.

We have shown in Section 6.3 that minimizing systemic risk requires an extremely powerful and active regulator, who would have to intervene three times more than is currently the case. In practice, this could only be achieved by increasing the regulator's discretionary power. Such a proactive position was also included in the proposed SAFE Banking Act of 2012, under which a maximum bank size relative to the system would be imposed. Notwithstanding a range of limited reweighting schemes, our results indicate that in our setup, systemic risk could not be reduced while maintaining the current size of the largest banks. However, in terms of optimal portfolio design, bringing their cumulative weight down from 90% to 50% yielded a significant improvement in systemic risk.

6.4.2 Does the result hold for a system where only listed banks are considered?

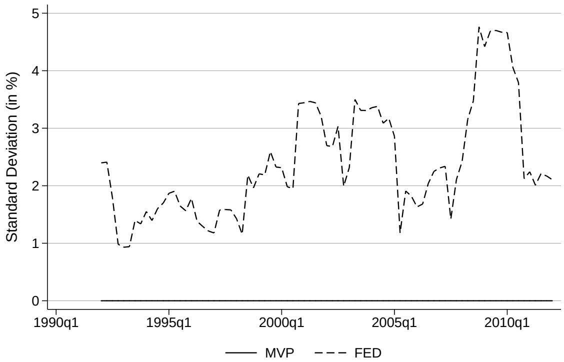
One assumption we have consistently made is that the regulator is able to move substantial amounts of assets from large banks to very small ones. However, small banks might not be able to sustain such an increase in assets in the first place. Moreover, previous research has shown that start-up banks only behave as mature banks after their first nine years of existence (see e.g. DeYoung and Hasan, 1998). Because these small banks might not be realistic investments for the supervisor, we select only those BHCs which have publicly traded equity, using the CRSP-FRB link provided by the Federal Reserve Bank of New York (2013). We thereby also remove those banks that do not file reports after 2006 due to the increase in the reporting threshold. The selected banks are considered to be the entire portfolio in which the supervisor can invest. The results in the Appendix show that removing these banks does not quantitatively or qualitatively change our results: It is possible to minimize portfolio risk by relocating assets from the largest listed banks to smaller listed ones. Moreover, in this new MVP, the actual largest 5% of banks would still receive a very low weight.[18]

Next, we use the CRSP-FRB link to obtain the market valuation of assets for listed BHCs by downloading equity prices as well as the number of outstanding shares from CRSP, and match them to the Call Reports. The quarterly market valuation of assets is obtained by adding the book value of the liabilities to the average market capitalization during that quarter. Compared with book value, which gives information on the past performance of a bank, the market valuation should indicate what market participants believe to be the value of the bank going forward, notwithstanding liquidity considerations during crises (Allen and Carletti, 2008). To perform our analysis on this subsample, we define returns as the quarter-to-quarter percentage changes in the market valuation of assets, and a banks' weight as the relative share in the portfolio. As in the baseline scenario, we obtain the MVP using an eight quarter moving window. If a

[18] A scenario in which the smallest 60% of banks were removed yielded similar results.

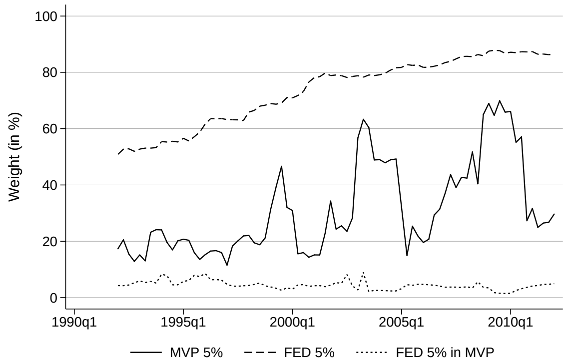
Figure 6.13: What role does inequality play in the risk/return trade-off when taking into account the market valuation of assets?

(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?

(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP: panel 6.13a and 6.13b display the difference in the risk and return of each portfolio. Panel 6.13c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

merger takes place during this time, the assets of the acquired bank are added to the acquiring bank before the merger takes place for the appropriate quarters, while the acquired bank is removed from the investment universe.[19] The results are reported in Figure 6.13.

Figure 6.13 shows spikes in the portfolio risk of the FED portfolio at the beginning of the 2000s and during the subprime crisis, whereas that of the MVP is essentially zero. While the returns of the MVP are slightly higher prior to 2005, they are similar

[19]Similar to the analyses on book data, banks that report a return below -100% or above +100% are not considered.

Table 6.4: Extensions and policy implications tests

	Issue	Policy test	Risk minimized? BHC	Risk minimized? COMM
(1)	High turnover	Equally Weighted Portfolio	✗	0/12
(2)	High turnover	Reweighting limited to 10% or 20% of assets[a]	✗	0/12
(3)	High turnover	Reweighting limited to mean/std. dev. of growth[a]	✗	0/12
(4)	High turnover	Largest 5% keep their weight[a]	✗	5/12
(5)	High turnover	Largest 5% are weighted between 50% and 60%	✓	12/12
(6)	Small/DeNovo banks	Only BHCs with publicly traded equity – book value	✓	n.a.
(7)	Small/DeNovo banks	Only BHCs with publicly traded equity – market value	✓	n.a.

The table presents a summary of the seven policy tests that were performed and how well they worked in reducing systemic risk on FED and FRD level. To see if a test is successful we define the ratio $(\sigma_{\text{FED}} - \sigma_{\text{MVP}})/\sigma_{\text{FED}}$, which measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements were possible. A test is regarded as successful when this ratio averages 0.9 or higher throughout the sample period.

[a] To avoid the no portfolio loss constraint driving these results, we also performed the tests without the no-loss constraint. The results however do not change.

to the FED portfolio thereafter. Regarding the weight distribution of both portfolios, we again see a steady increase in the weights of the largest banks in the FED portfolio. The MVP based on market valuation shows a higher concentration than before, as the largest banks on average are assigned 33% of the assets compared with 13% in the baseline scenario. Similar to the baseline scenario, concentration in the MVP peaks in the subprime crisis when the largest banks have a weight of 69%. However, it appears that the largest banks in the actual portfolio are still overweighted, as they see their weight reduced to an average of 4.4% in the MVP. Table 6.4 summarizes the results of all scenarios in this section.

6.5 Conclusion

The last two decades have seen a major wave of consolidation and concentration of assets in the banking industry. In the same period, the sector has experienced two major crises with a significant impact on the real economy, of which the subprime crisis had global repercussions. As a consequence of recent bailouts and government-forced sales, the sector is now even more concentrated than before the crises. In the light of moral hazard and Too-Big-To-Fail banks, we have investigated how the high concentration in the industry impacts systemic risk.

In the absence of counterfactuals, we consider a thought experiment in which we view the supervisor as a constrained investor in a portfolio of banks. As profit maximization by banks is inherently risky, but is not perfectly correlated with that of other banks, the portfolio of the supervisor will have a better risk-return profile. By applying elements of Modern Portfolio Theory, we derive a hypothetical distribution of weights that the supervisor should have held to arrive at the minimum variance portfolio in order to give us insights into the optimal design of the banking system.

Our results consistently show that the hypothetical minimum variance portfolio had a lower risk than the actual portfolio, achieved by reducing the level of concentration in the portfolio. Moreover, it was not necessary to sacrifice returns in order to achieve this lower risk. The minimum variance portfolio favors more traditional banks as measured by the non-interest income/total income ratio and balance sheet items such as loans, trading assets and deposits. In contrast, an equally-weighted portfolio would perform similarly to the actual, concentrated, system. These findings are robust to different starting values, time windows, covariance matrices and the exclusion of the smallest banks.

However, to achieve lower risk, the supervisor would have to adjust weights in each quarter, leading to a portfolio turnover three times higher than that of the real portfolio. Since this might not be possible or even desirable within the current regulatory framework, we tested several alternatives involving limited reweighting which were largely unsuccessful. These findings indicate that in times of crisis, an increase in systemic risk was unavoidable while keeping the concentration at current levels. Nonetheless, our analysis did show - ceteris paribus - that when the weight of the largest banks was kept at a sufficiently low level, systemic risk was reduced significantly in the hypothetical minimum variance portfolio.

The policy implications flowing from these findings are that supervisors should seriously consider the effects of concentration on systemic risk. A reduction in disparity of size could create a more competitive environment, similar to provisions of the proposed Safe, Accountable, Fair & Efficient (SAFE) Banking Act of 2012, which would limit individual banks' funding strategies to 10% of the total industry. Other measures could include imposing higher equity capital demands for the large banks as are currently being implemented, or separating investment from retail banking as proposed by Paul Volcker, and the Vickers and Liikanen reports in Europe. Forcing large banks to hold more capital could lead to a relative reduction in their size only if their assets are 'redistributed' to smaller banks in order to maintain a safer and more competitive environment. Our findings show that we should not only consider the size of each bank individually, but also consider each bank's size with respect to the whole system. However, given the data at hand which mirrors the current regulatory set-up, the results presented here can only be interpreted as a partial equilibrium effect as we can not consider the shifting of bank activities abroad or into the shadow banking realm.

This Appendix contains results that were omitted from the body of the paper for brevity. In Section 6.A, we report the results of the baseline analysis for each of the separate Federal Reserve Districts (FRDs). Section 6.B shows results for several robustness tests, while Section 6.C shows results for the extensions and policy implications tests.

Appendix 6.A Extensive results: Baseline specification

Tables 6.A.1 and 6.A.2 show the differences in the risk-return trade-off and distribution of weights for the FRD portfolios and their respective MVPs. We show the results for the entire sample (1984Q4 - 2010Q4), as well as different subsamples (1984Q4 - 1993Q4, 1994Q1 - 2006Q4 and 2007Q1 - 2010Q4). The numbers reported are averages during the relevant time span and Kolmogorov-Smirnov tests are performed to see if the distributions in the FRDs and MVPs are different from each other. In the second panel of Table 6.A.1, we report the relative difference in portfolio risk between the MVPs and FRD portfolios. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements were possible.

The tables show mixed results for the differences in return over the whole sample: some FRDs outperform their MVPs, whereas others exhibit lower returns. One interesting fact is that in the boom period of 1994Q1 - 2006Q4, we find that all FRDs outperform their MVP counterparts in terms of returns. Regarding size disparity, the MVPs constantly have a much lower level of concentration compared to their FRD portfolios, their difference ranging on average between 44% and 78% throughout the sample period.

Table 6.A.1: Portfolio optimization baseline - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{MVP} - r_{FRD}$				
Boston	0.032***	0.131***	−0.037**	0.026
New York	0.044***	0.167***	−0.016***	−0.045*
Philadelphia	−0.105***	0.036***	−0.271***	0.107**
Cleveland	0.011***	0.036***	−0.026***	0.073**
Richmond	−0.023***	0.036***	−0.050***	−0.075*
Atlanta	−0.028***	−0.042**	−0.047***	0.066***
Chicago	0.018***	0.042***	−0.017**	0.079***
St. Louis	−0.018***	−0.020***	−0.030***	0.023
Minneapolis	−0.123***	−0.091***	−0.150***	−0.113**
Kansas City	−0.047***	−0.063***	−0.044***	−0.019
Dallas	0.012***	0.070**	−0.019***	−0.019
San Francisco	−0.081***	−0.075**	−0.102***	−0.030*
$(\sigma_{FRD} - \sigma_{MVP})/\sigma_{FRD}$				
Boston	1.000***	1.000***	1.000***	1.000***
New York	1.000***	1.000***	1.000***	1.000***
Philadelphia	1.000***	1.000***	1.000***	1.000***
Cleveland	1.000***	1.000***	1.000***	1.000***
Richmond	1.000***	1.000***	1.000***	1.000***
Atlanta	1.000***	1.000***	1.000***	1.000***
Chicago	1.000***	1.000***	1.000***	1.000***
St. Louis	1.000***	1.000***	1.000***	1.000***
Minneapolis	1.000***	1.000***	1.000***	1.000***
Kansas City	1.000***	1.000***	1.000***	1.000***
Dallas	1.000***	1.000***	1.000***	1.000***
San Francisco	1.000***	1.000***	1.000***	1.000***

This table shows the results of the baseline specification for the analysis on commercial bank level. $r_{MVP} - r_{FRD}$measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{FRD} - \sigma_{MVP})/\sigma_{FRD}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Table 6.A.2: Portfolio optimization baseline - Weights

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−68.8***	−59.7***	−74.0***	−73.3***
New York	−76.1***	−75.3***	−80.4***	−64.2***
Philadelphia	−58.0***	−45.2***	−60.6***	−79.3***
Cleveland	−75.6***	−62.3***	−79.8***	−92.6***
Richmond	−78.0***	−66.8***	−82.4***	−89.3***
Atlanta	−64.5***	−59.8***	−67.2***	−66.3***
Chicago	−62.2***	−55.1***	−67.5***	−61.7***
St. Louis	−48.5***	−43.6***	−52.4***	−47.1***
Minneapolis	−62.4***	−50.5***	−63.9***	−85.0***
Kansas City	−43.8***	−38.3***	−47.3***	−45.2***
Dallas	−52.3***	−51.3***	−54.0***	−49.1***
San Francisco	−71.1***	−75.2***	−67.8***	−72.3***
MVP 5% - FRD 5%				
Boston	−54.2***	−42.2***	−63.4***	−52.0***
New York	−65.7***	−61.4***	−73.5***	−49.9***
Philadelphia	−48.5***	−32.5***	−55.1***	−64.1***
Cleveland	−68.1***	−52.6***	−74.9***	−82.1***
Richmond	−68.0***	−53.7***	−74.8***	−79.1***
Atlanta	−55.3***	−48.0***	−62.2***	−49.3***
Chicago	−55.8***	−48.2***	−62.4***	−52.0***
St. Louis	−38.4***	−33.7***	−42.7***	−35.1***
Minneapolis	−52.6***	−40.2***	−55.2***	−73.0***
Kansas City	−35.4***	−31.9***	−37.9***	−35.7***
Dallas	−44.0***	−42.3***	−46.9***	−38.5***
San Francisco	−62.4***	−65.3***	−62.6***	−55.2***

This table shows the results of the baseline specification for the analysis on commercial bank level. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Appendix 6.B Extensive results: Robustness tests

In this section, we revisit several assumptions in the baseline specification and, where possible, relax them. In section 6.B.1, we first explore the assumption that the sample covariance matrix does not change when banks face a large increase/decrease in total assets. Second, in section 6.B.2, we use an alternative estimator of the covariance matrix which reduces the noise in a small T and large N setting employed in our baseline specification. Third, section 6.B.3 examines whether the baseline findings are a result of the starting weights, before confirming in section 6.B.4 that the results also hold for a different length window on which the covariance matrix is estimated.

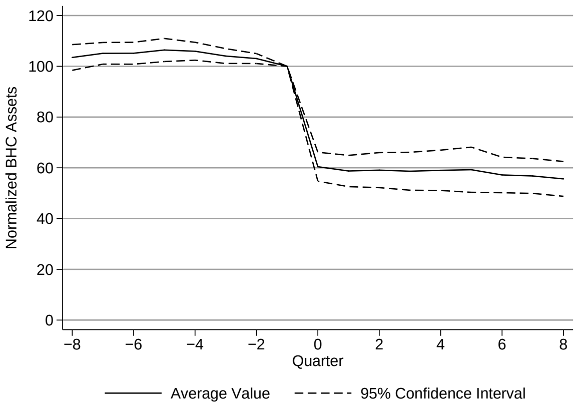
6.B.1 Identifying Jumps in Assets of BHCs for Portfolio Optimization under a Dynamic Covariance and Return Matrix

In the baseline specification, the sample covariance matrix was assumed to be constant, such that a large change in the size of the bank would not have changed its return structure. Although we have shown that large and small banks share a common support in this structure, here we identify cases in which banks have seen a large increase/decrease in their total assets and analyze how the elements of its covariance matrix changed.

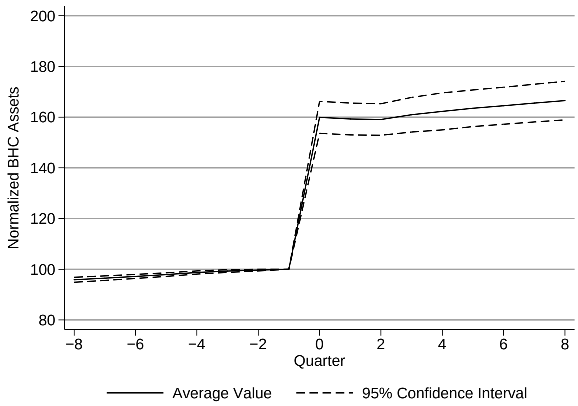
A jump is defined as an increase/decrease of bank assets of 25% or greater from one quarter to the next, provided that the preceding and following 8 quarters did not show jumps larger than 10% in each of the quarters, nor a cumulative change in the preceding and following 8 quarters of 25%. These last two conditions are imposed to make sure that bank size before and after the jump was relatively stable and that the change in the elements of the covariance matrix can be chiefly attributed to the one-time jump. We find 15 negative jumps and 287 positive jumps during the sample period. Figure 6.B.1 depicts the movement in assets before and after the jump, where the total assets are normalized to 100 at the quarter prior to the jump. Banks that experience a negative jump lose on average 40% of their assets, whereas banks experiencing a positive jump gain on average 60%.

Figure 6.B.1: Jump in bank assets

(a) Negative Jump



(b) Positive Jump



The figure displays the average jump in total bank assets for the selected bank holding companies, as well as its 95% confidence interval. The negative jump in assets is based on 15 BHCs, whereas the positive jump is based on 287 BHCs. Jumps were defined as a decrease (increase) of bank assets of 25% or greater from one quarter to next, provided that the preceding and following 8 quarters did not show jumps larger than 10% in each of the quarters, nor a cumulative change in the preceding and following 8 quarters of 25%. Assets are normalized to 100 in the quarter before the jump.

Table 6.B.1: Changes in the return and covariance matrices due to jumps

		ROA		Variance ROA		Covariance ROA	
		Pre-Jump	Post-Jump	Pre-Jump	Post-Jump	Pre-Jump	Post-Jump
Positive Jump	Value	0.341	0.259	0.023	0.111	0.002	0.005
	Average Difference	-0.082		0.088		0.004	
	Percentage Difference	-23.97%		379.70%		233.70%	
	P-value T-test	0.000		0.007		0.010	
Negative Jump	Value	-0.142	0.213	0.281	3.249	0.007	-0.015
	Average Difference	0.355		2.967		-0.022	
	Percentage Difference	249.52%		1054.64%		-324.52%	
	P-value T-test	0.110		0.359		0.293	

The table shows the average differences in ROA, its variance and its average pairwise covariance pre- and post-jump for both negative jumps and positive jumps.

The average changes to the return and covariance matrices are displayed in Table 6.B.1. T-tests show that banks experiencing a positive jump in assets have a statistically significant lower return-on-assets 8 quarters following the jump and a higher average variance and covariance of these returns. On the other hand, banks experiencing a negative jump do not see statistically significant changes in their average variance and covariance, and only a marginally significant increase in their return-on-assets after the jump.

Using the statistically significant changes for positive and negative shocks, we interpolate the effect to the sample covariance matrix and the return matrix, such that these matrices change dynamically with the weights received by the banks. As the change in return-on-assets for the banks experiencing a negative jump is marginally significant at 11%, we also regard this change as significant. During the iterative process, the covariance matrix is updated based on the proposed weights and the actual weights. For example, if a bank has a weight of 5% and the outcome of the iteration is that it should have a weight of 7.5%, i.e. an increase of 50%, its return-on-assets would decrease by $\frac{0.50}{0.60} \times -23.97\% = -19.975\%$, its variance would increase by $\frac{0.50}{0.60} \times 379.70\% = 316.416\%$ and every covariance term would increase by $\frac{0.50}{0.60} \times 233.70\% = 194.75\%$. If, based on the new return and covariance matrices, in the second iteration the proposed weight is 4%, i.e. a 20% decrease from its original size, only its return-on-assets would increase by $\frac{0.20}{0.40} \times 249.52\% = 124.76\%$, while nothing would happen to its original variance and covariance terms as these changes are not statistically significant. Since we do not have data on jumps beyond -40% and +60%, any proposed weight change beyond these thresholds will lead to an adjustment of the return and covariance matrices as if they were -40% or +60%.

In the first robustness test, we adjust only the return matrix and then estimate a new sample covariance matrix. In the second test, we adjust both the return matrix and the covariance matrix with which the portfolio risk will be minimized. The results are shown in the paper in Figure 8.

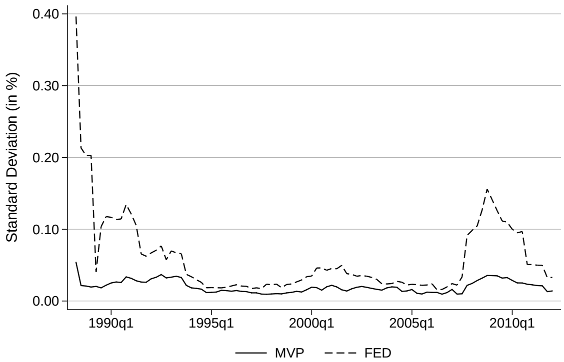
6.B.2 Portfolio Optimization using the Ledoit and Wolf (2003, 2004) Covariance Matrix

The sample covariance matrix with which portfolio optimization is applied, relies on a large N and small T setting, which has been shown to introduce a considerable amount of noise in this estimation. Ledoit and Wolf (2003, 2004) proposed a shrinkage based estimator of the covariance matrix to reduce the noise. We use their proposed estimator of the covariance matrix instead of its sample equivalent and re-run the analysis. Figure 6.B.2 show the results for the analysis on BHC level, and Tables 6.B.2 and 6.B.3 do the same for the analysis on commercial bank level.

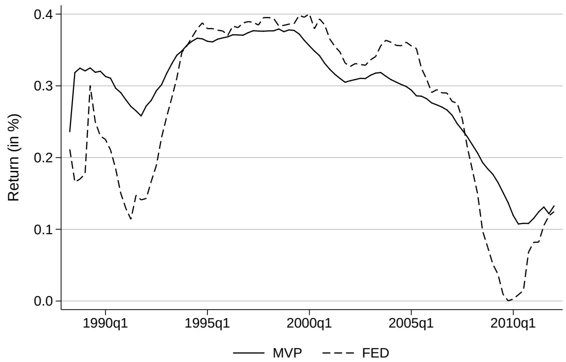
Overall, we find that the standard deviation of the MVP is now higher compared to the baseline specification, although still lower than that of the FED portfolio. The ratio $(\sigma_{\text{FED}} - \sigma_{\text{MVP}})/\sigma_{\text{FED}}$ is on average only 0.54, indicating that 54% of the portfolio risk was minimized. For the commercial banks, this ratio ranges between 0.40 and 0.88. The portfolio returns, meanwhile, are at a similar level. The levels of concentration in the MVP are basically unchanged, as the largest banks still have a weight of on average 10%, while the largest banks of the FED portfolio have a weight in the MVP of 5%. For the FRD portfolios, on average 60% of portfolio risk is eliminated, and the return of the MVPs is higher. The weight of the largest 5% of banks in the MVP is on average 47% lower than in the FRD portfolio, while the largest banks in the FRD are reduced by 70%. Both of these results are similar to the baseline specification.

Figure 6.B.2: Portfolio optimization using the Ledoit and Wolf (2003, 2004) covariance matrix

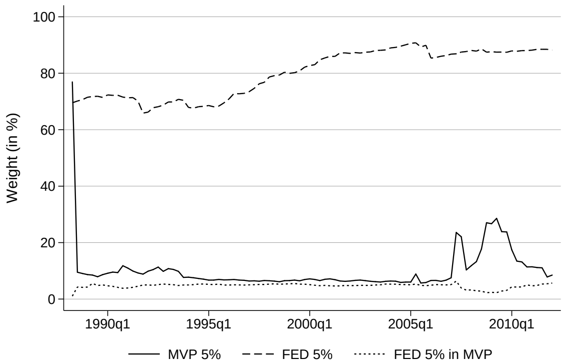
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP using the Ledoit and Wolf (2003, 2004) shrinkage estimator of the covariance matrix: panel 6.B.2a and 6.B.2b display the difference in the risk and return of each portfolio. Panel 6.B.2c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.B.2: Portfolio optimization using the Ledoit and Wolf (2003, 2004) covariance matrix - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\text{MVP}} - r_{\text{FRD}}$				
Boston	0.024***	0.124***	−0.055***	0.047*
New York	0.089***	0.234***	0.008***	0.016**
Philadelphia	−0.103***	0.032***	−0.288***	0.189**
Cleveland	−0.016***	0.060***	−0.079***	0.017**
Richmond	0.020***	0.090***	−0.028***	0.012*
Atlanta	0.032***	0.078***	−0.036***	0.149***
Chicago	0.035***	0.073***	−0.011*	0.098***
St. Louis	0.010***	0.046***	−0.035***	0.073***
Minneapolis	−0.099***	−0.000***	−0.182***	−0.055*
Kansas City	0.022***	0.083***	−0.023***	0.029*
Dallas	0.100***	0.262***	0.006***	0.031**
San Francisco	−0.003***	0.105***	−0.100***	0.064***
$(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$				
Boston	0.847***	0.787***	0.885***	0.861***
New York	0.801***	0.737***	0.806***	0.929***
Philadelphia	0.880***	0.810***	0.902***	0.970***
Cleveland	0.729***	0.664***	0.752***	0.808***
Richmond	0.681***	0.525***	0.778***	0.730***
Atlanta	0.402***	0.358***	0.342***	0.699***
Chicago	0.486***	0.495***	0.429***	0.650***
St. Louis	0.394***	0.336***	0.351***	0.666***
Minneapolis	0.587***	0.569***	0.546***	0.762***
Kansas City	0.316***	0.380***	0.224***	0.465***
Dallas	0.501***	0.634***	0.385***	0.572***
San Francisco	0.712***	0.687***	0.726***	0.727***

This table shows the results of the Ledoit and Wolf (2003, 2004) shrinkage based estimator of the covariance matrix for the analysis on commercial bank level. $r_{\text{MVP}} - r_{\text{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6.B.3: Portfolio optimization using the Ledoit and Wolf (2003, 2004) covariance matrix - Weights

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−63.7***	−53.1***	−66.4***	−79.2***
New York	−74.4***	−70.3***	−79.9***	−65.9***
Philadelphia	−58.9***	−43.9***	−62.3***	−82.6***
Cleveland	−74.8***	−58.6***	−80.4***	−94.0***
Richmond	−76.0***	−60.8***	−82.3***	−90.4***
Atlanta	−61.1***	−56.6***	−62.5***	−67.1***
Chicago	−59.9***	−48.4***	−67.5***	−61.8***
St. Louis	−47.1***	−39.5***	−52.4***	−47.1***
Minneapolis	−61.4***	−47.8***	−64.0***	−84.9***
Kansas City	−42.1***	−36.3***	−45.4***	−45.0***
Dallas	−50.8***	−49.5***	−52.5***	−48.3***
San Francisco	−67.1***	−65.7***	−66.8***	−71.1***
MVP 5% - FRD 5%				
Boston	−26.0***	−9.7***	−36.2***	−30.2***
New York	−47.9***	−29.9***	−65.4***	−32.9***
Philadelphia	−30.7***	−7.4***	−44.3***	−40.3***
Cleveland	−60.2***	−40.5***	−68.3***	−79.5***
Richmond	−62.0***	−44.5***	−73.9***	−63.4***
Atlanta	−53.9***	−47.2***	−59.8***	−49.9***
Chicago	−57.8***	−46.8***	−65.9***	−57.0***
St. Louis	−41.5***	−31.5***	−48.9***	−40.6***
Minneapolis	−53.8***	−38.2***	−58.4***	−75.1***
Kansas City	−37.4***	−29.1***	−42.5***	−40.1***
Dallas	−41.2***	−33.8***	−46.2***	−42.3***
San Francisco	−50.9***	−49.4***	−59.2***	−27.5***

This table shows the results of the Ledoit and Wolf (2003, 2004) shrinkage based estimator of the covariance matrix for the analysis on commercial bank level. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

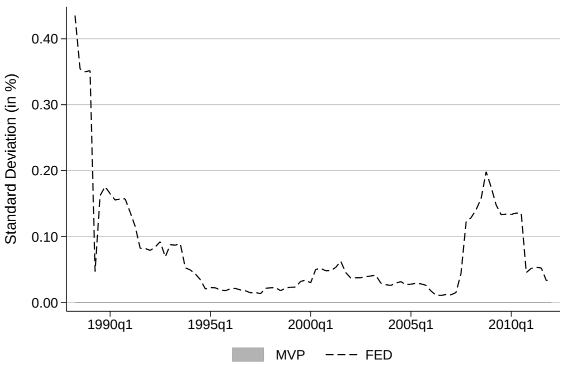
6.B.3 Equal and Random Starting Weights

Since our baseline specification deals with many banks, the minimization of the portfolio risk is likely to be a complex, highly nonlinear problem comprising multiple minima/solutions. The starting values for the optimization are chosen to be the actual weights in the portfolio, and can have a large impact on whether a global or local minimum is found as well as in what direction the distribution of weights will move. To account for this possible bias, we run two robustness tests. First, we run, for BHC data only, 100 repetitions per quarter using randomized starting values. The random starting values are drawn from a half-normal distribution and then divided by its sum, such that they add up to 1. In the second test, we choose, for the commercial bank data, as starting value an equally weighted portfolio.

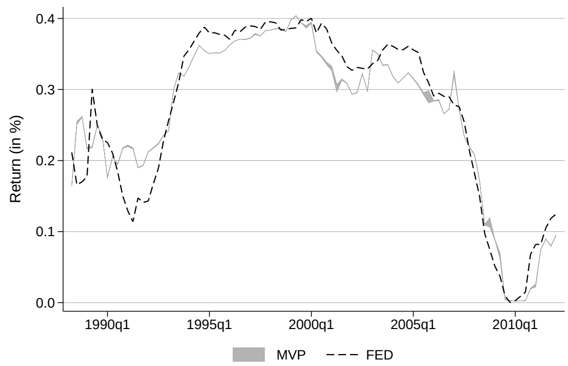
Figure 6.B.3 shows the results for the random starting values, while Figure 6.B.4 Tables 6.B.4 and 6.B.5 show the results for the equal starting weights. Figure 6.B.3 shows the minimum and maximum portfolio risk, return and concentration found in each period and we can see that there is hardly any variation in the outcomes. The optimization using equal starting weights show almost identical results to those of the baseline specification. Overall, we take this as evidence that we approach a global optimum in the baseline specification.

Figure 6.B.3: Do different starting weights values matter in the optimization?

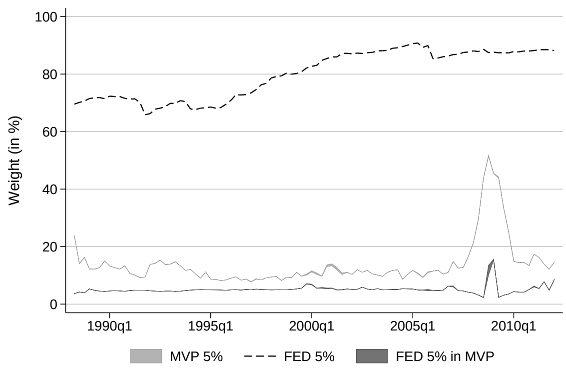
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



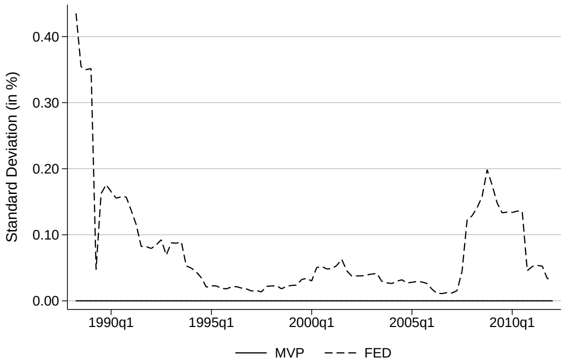
(c) How unequal is the MVP compared to the FED portfolio?



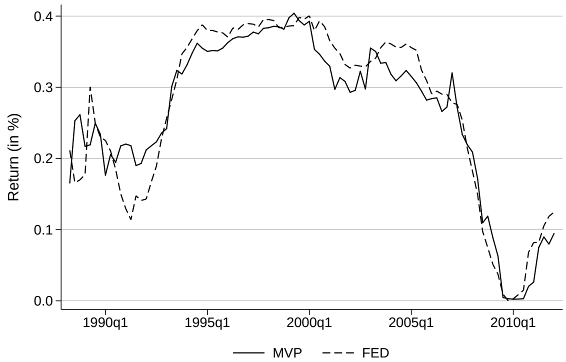
The figure presents the robustness test in which 100 random starting values were chosen per quarter, to test whether using the actual weights as starting values drives the results. The lower and upper bound are displayed to show that starting values do not matter.

Figure 6.B.4: Do different starting weights values matter in the optimization?

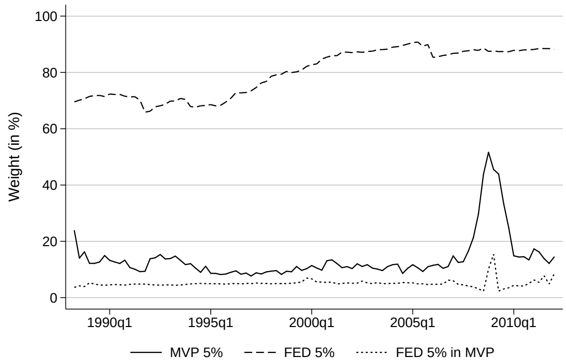
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The figure presents the robustness test in which an equally weighted portfolio was chosen as starting weights, instead of the actual weights.

Table 6.B.4: Portfolio optimization using equal starting weights - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\mathrm{MVP}} - r_{\mathrm{FRD}}$				
Boston	0.032***	0.131***	−0.037***	0.027
New York	0.044***	0.167***	−0.016***	−0.046*
Philadelphia	−0.105***	0.036***	−0.271***	0.107**
Cleveland	0.011***	0.036***	−0.025***	0.074**
Richmond	−0.023***	0.037***	−0.050***	−0.075*
Atlanta	−0.028***	−0.041*	−0.047***	0.065***
Chicago	0.018***	0.042***	−0.017**	0.079***
St. Louis	−0.019***	−0.020***	−0.032***	0.023
Minneapolis	−0.123***	−0.091***	−0.150***	−0.114**
Kansas City	−0.046***	−0.063***	−0.043***	−0.020
Dallas	0.012***	0.070**	−0.019***	−0.019
San Francisco	−0.081***	−0.075**	−0.100***	−0.030*
$(\sigma_{\mathrm{FRD}} - \sigma_{\mathrm{MVP}})/\sigma_{\mathrm{FRD}}$				
Boston	1.000***	1.000***	1.000***	1.000***
New York	1.000***	1.000***	1.000***	1.000***
Philadelphia	1.000***	1.000***	1.000***	1.000***
Cleveland	1.000***	1.000***	1.000***	1.000***
Richmond	1.000***	1.000***	1.000***	1.000***
Atlanta	1.000***	1.000***	1.000***	1.000***
Chicago	1.000***	1.000***	1.000***	1.000***
St. Louis	1.000***	1.000***	1.000***	1.000***
Minneapolis	1.000***	1.000***	1.000***	1.000***
Kansas City	1.000***	1.000***	1.000***	1.000***
Dallas	1.000***	1.000***	1.000***	1.000***
San Francisco	1.000***	1.000***	1.000***	1.000***

This table shows the results of the analysis on commercial bank level when using equal starting weights. $r_{\mathrm{MVP}} - r_{\mathrm{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\mathrm{FRD}} - \sigma_{\mathrm{MVP}})/\sigma_{\mathrm{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Table 6.B.5: Portfolio optimization using equal starting weights - Weights

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−68.8***	−59.7***	−74.0***	−73.3***
New York	−76.1***	−75.3***	−80.4***	−64.2***
Philadelphia	−58.0***	−45.2***	−60.6***	−79.4***
Cleveland	−75.6***	−62.3***	−79.8***	−92.6***
Richmond	−78.0***	−66.8***	−82.4***	−89.3***
Atlanta	−64.5***	−59.8***	−67.2***	−66.4***
Chicago	−62.2***	−55.1***	−67.4***	−61.7***
St. Louis	−48.5***	−43.6***	−52.4***	−47.1***
Minneapolis	−62.4***	−50.5***	−63.9***	−85.0***
Kansas City	−43.8***	−38.3***	−47.3***	−45.2***
Dallas	−52.3***	−51.3***	−54.0***	−49.1***
San Francisco	−71.1***	−75.2***	−67.8***	−72.3***
MVP 5% - FRD 5%				
Boston	−54.2***	−42.2***	−63.4***	−51.6***
New York	−65.7***	−61.4***	−73.5***	−49.9***
Philadelphia	−48.5***	−32.5***	−55.1***	−64.1***
Cleveland	−68.1***	−52.6***	−74.9***	−82.1***
Richmond	−68.0***	−53.7***	−74.8***	−79.1***
Atlanta	−55.3***	−48.0***	−62.2***	−49.3***
Chicago	−55.8***	−48.2***	−62.4***	−52.0***
St. Louis	−38.3***	−33.7***	−42.6***	−35.1***
Minneapolis	−52.6***	−40.2***	−55.2***	−72.9***
Kansas City	−35.4***	−31.9***	−37.8***	−35.7***
Dallas	−44.0***	−42.3***	−46.9***	−38.5***
San Francisco	−62.4***	−65.3***	−62.6***	−55.2***

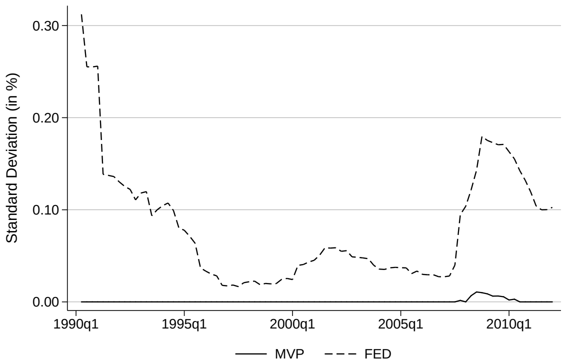
This table shows the results of the analysis on commercial bank level when using equal starting weights. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

6.B.4 Portfolio Optimization using a 16 Quarter Window

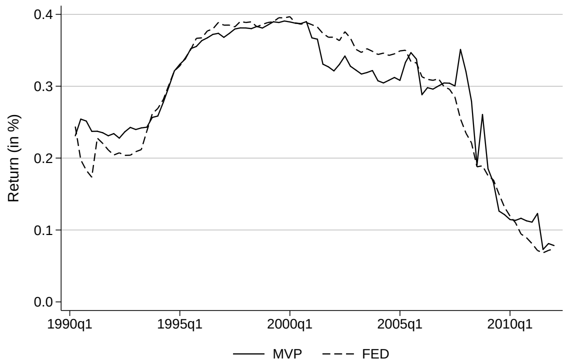
Finally, we explore if the choice of an 8 quarter window on which the covariance matrices are estimated matters for the optimization. In order to see if our results are robust, we run the analysis using a 16 quarter window. The results for the analysis on BHC level is shown in Figure 6.B.5, whereas the results for the analysis on commercial bank level are shown in Tables 6.B.6 and 6.B.7, with being quantitatively and qualitatively similar to the baseline specification.

Figure 6.B.5: Robustness - 16 quarter analysis

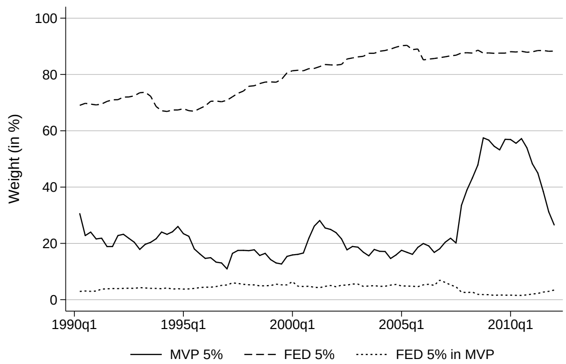
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP when using 16 quarters to estimate the sample covariance matrix: panel 6.B.5a and 6.B.5b display the difference in the risk and return of each portfolio. Panel 6.B.5c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.B.6: Portfolio optimization using a 16 quarter window - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\text{MVP}} - r_{\text{FRD}}$				
Boston	0.015***	0.140***	−0.045***	−0.018*
New York	0.048***	0.202***	−0.003***	−0.063***
Philadelphia	−0.143***	0.081***	−0.269***	−0.139***
Cleveland	−0.025***	0.048***	−0.066***	−0.026
Richmond	−0.010***	0.050***	−0.025***	−0.073*
Atlanta	−0.036***	−0.060***	−0.044***	0.032
Chicago	0.013***	0.042***	−0.009*	0.036*
St. Louis	−0.026***	−0.023***	−0.037***	0.005
Minneapolis	−0.136***	−0.060***	−0.176***	−0.145***
Kansas City	−0.054***	−0.087***	−0.042***	−0.035**
Dallas	−0.002***	0.056***	−0.024***	−0.035*
San Francisco	−0.068***	−0.039***	−0.094***	−0.037
$(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$				
Boston	0.993***	0.996***	1.000***	0.963***
New York	1.000***	1.000***	1.000***	1.000***
Philadelphia	1.000***	1.000***	1.000***	1.000***
Cleveland	1.000***	1.000***	0.999***	1.000***
Richmond	1.000***	1.000***	1.000***	0.997***
Atlanta	0.999***	1.000***	1.000***	0.995***
Chicago	1.000***	1.000***	1.000***	1.000***
St. Louis	1.000***	1.000***	1.000***	1.000***
Minneapolis	1.000***	1.000***	1.000***	1.000***
Kansas City	1.000***	1.000***	1.000***	1.000***
Dallas	1.000***	1.000***	1.000***	1.000***
San Francisco	1.000***	1.000***	1.000***	1.000***

This table shows the results of the analysis on commercial bank level when using 16 quarters to estimate the sample covariance matrix. $r_{\text{MVP}} - r_{\text{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Table 6.B.7: Portfolio optimization using a 16 quarter window - Weights

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−69.5***	−60.5***	−72.7***	−75.6***
New York	−75.8***	−75.1***	−80.1***	−62.9***
Philadelphia	−58.8***	−44.6***	−60.8***	−78.1***
Cleveland	−75.8***	−62.0***	−78.6***	−91.6***
Richmond	−79.0***	−66.4***	−82.8***	−89.3***
Atlanta	−65.7***	−61.7***	−67.7***	−66.4***
Chicago	−63.0***	−55.7***	−67.3***	−62.0***
St. Louis	−49.0***	−44.5***	−51.9***	−47.5***
Minneapolis	−61.7***	−51.7***	−61.2***	−81.8***
Kansas City	−44.1***	−38.2***	−46.9***	−46.0***
Dallas	−51.0***	−48.1***	−54.1***	−46.3***
San Francisco	−66.1***	−74.1***	−67.3***	−47.9***
MVP 5% - FRD 5%				
Boston	−35.5***	−14.5***	−48.6***	−31.4***
New York	−53.3***	−41.5***	−65.3***	−36.0***
Philadelphia	−35.9***	−18.1***	−44.8***	−39.6***
Cleveland	−57.8***	−42.4***	−63.8***	−66.3***
Richmond	−54.8***	−39.8***	−64.3***	−51.0***
Atlanta	−44.8***	−37.9***	−54.7***	−25.4***
Chicago	−49.7***	−43.3***	−56.5***	−39.1***
St. Louis	−27.6***	−25.8***	−30.4***	−22.0***
Minneapolis	−42.8***	−32.8***	−45.1***	−53.0***
Kansas City	−27.6***	−26.0***	−30.0***	−22.5***
Dallas	−33.0***	−28.5***	−38.2***	−24.1***
San Francisco	−49.7***	−54.3***	−57.0***	−17.4***

This table shows the results of the analysis on commercial bank level when using 16 quarters to estimate the sample covariance matrix. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Appendix 6.C Extensive results: Extensions and policy implications

In this final section of the Appendix, we examine alternative scenarios which approximate better the current regulatory framework. In section 6.C.1 we examine several scenarios that would limit the amount of reweighting a supervisor could perform when determining its MVP. Finally, in section 6.C.2 we repeat the analysis on a more realistic sample of banks in which the supervisor could invest.

6.C.1 Portfolio Optimization under Limited reweighting

Given the large amount of turnover in the baseline specification, a more realistic scenario would be that the supervisor is able to change the size of a bank within certain limitations. We explore several of these scenario's here, starting with a static constraint that the bank supervisor can add or subtract a certain percentage of an individual institution. Figure 6.C.1 and Tables 6.C.1-6.C.2 show the results for the scenario when supervisors can not change banks' total assets more than 10%, and Figure 6.C.2 and Tables 6.C.3-6.C.4 do the same for 20%. For both BHCs and commercial banks, any improvement is hardly noticeable for the 10%-scenario, although the portfolio risk is somewhat lower for the 20%-scenario. Neither, however, match up to the baseline scenario and still exhibit the twin peaks in portfolio risk during the two crisis periods.

Another scenario is that the percentage with which the supervisor can reweigh his individual holdings depends on the business cycle. We construct a measure of asset growth in the quarter before the supervisor can set her weights, and allow reweighting equal to at most the average growth and the standard deviation of this growth: when average growth is larger, or more dispersed, this allows for more reweighting by the supervisor in order to reduce systemic risk. The measure, which acts as a time-varying constraint, is shown in Figure 6.C.3 using the data on asset growth for the BHCs. Indeed we see that the measures indeed exhibit time-variation, especially the standard deviation of the growth. In fact, the latter is even above 1 in 1997Q4, meaning that based on this constraint alone, it would be possible for the supervisor to short her banks. Since this is not possible, we cut the value off at 1 and thus end up in a situation of unlimited reweighting similar to the baseline specification for this quarter.

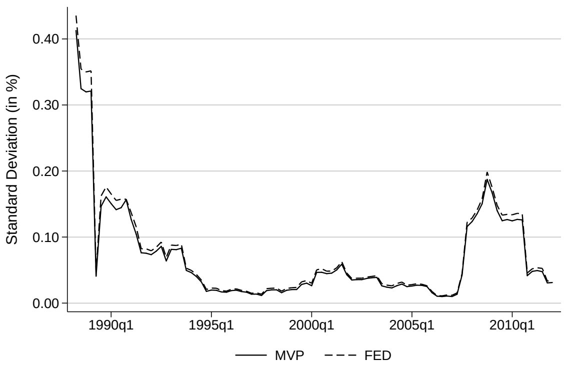
Figures 6.C.4 and Tables 6.C.5-6.C.6 display the results where the maximum amount of reweighting equals the mean of the asset growth in the previous quarter, whereas Figures 6.C.5 and Tables 6.C.7-6.C.8 do the same for the standard deviation of the growth. Outside of 1997Q4, which equals the baseline specification, there are no real improvements for both BHCs and commercial banks.

So far the supervisor has only been able to reweigh based on the weight that the bank had in the actual portfolio. In the next test we allow for dynamic reweighting where the supervisor can change the size of a bank with a percentage of its size in the previous quarter's MVP. In the first quarter that a bank is eligible for the portfolio, it will enter with its actual size. However, once it has been given an MVP weight (which lies in a range around its original size), the supervisor will keep on minimizing with the constraint on the size the bank had in the MVP rather than the actual portfolio. Since this is computationally more intensive, we perform this test only on BHC level. Figure 6.C.6 shows the results for a dynamic reweighting of a maximum of 10%, while Figure 6.C.7 shows the results for 20%. We can see no real improvement, although portfolio risk is again reduce more in the 20%-scenario compared to the 10%-scenario.

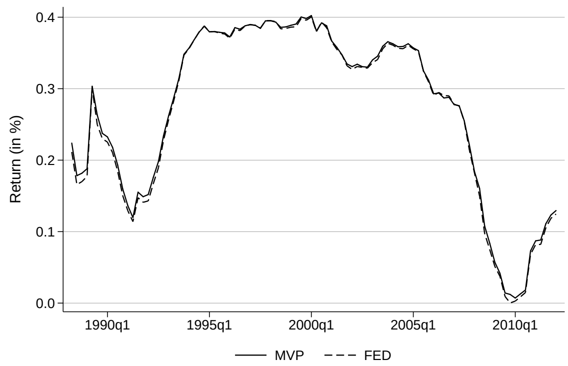
Given the inability to reduce portfolio risk when constraining the reweighting of individual banks, we finally explore some scenarios in which we constrain reweighting of the largest 5% of banks. Specifically, we add a constraint that the largest 5% of banks need to retain their current individual size, and allow unlimited reweighting of the remaining banks while still adhering to the no-shorting and no-loss constraint. The results for the commercial bank level are shown in Tables 6.C.9 and 6.C.10. There are 3 districts which were able to reduce systemic risk to a minimum in each of the three sub-samples. Interestingly, a common feature they share is that they have the lowest levels of concentration among all districts. Since the weight of these largest banks in the three successful districts never exceeds 60%, this leads us to a final test using limited reweighting: is it possible to reduce portfolio risk while keeping the cumulative weight of the largest 5% of banks between 50% and 60%? The results for the commercial banks are reported in Tables 6.C.11 and 6.C.12. We observe that under this limited reweighting scheme, it is possible to effectively eliminate risk while maintaining similar returns.

Figure 6.C.1: Portfolio optimization using limited reweighting - 10%

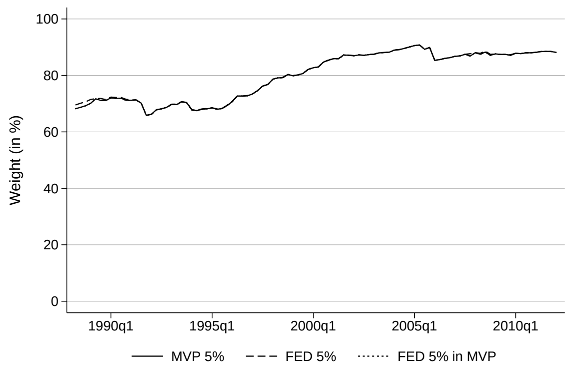
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The graph shows the comparison between the FED portfolio and the hypothetical MVP when constraining individual reweighting to 10% of the original bank size: panel 6.C.1a and 6.C.1b display the difference in the risk and return of each portfolio. Panel 6.C.1c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.C.1: Portfolio optimization using limited reweighting of 10% - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\text{MVP}} - r_{\text{FRD}}$				
Boston	0.004***	0.008	0.002	−0.000
New York	0.004***	0.006	0.004	0.001
Philadelphia	0.000***	0.005	−0.008	0.018
Cleveland	0.003***	0.006	0.002	0.002
Richmond	0.003***	0.005	0.001	0.002
Atlanta	0.002***	0.004	−0.001	0.010
Chicago	0.004***	0.006	0.001	0.009
St. Louis	0.002***	0.002	0.000	0.006
Minneapolis	−0.004***	0.001	−0.008	−0.003
Kansas City	0.002***	0.005	−0.000	0.004
Dallas	0.005***	0.012	0.000	0.003
San Francisco	0.002***	0.007	−0.003	0.004
$(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$				
Boston	0.070***	0.070	0.074	0.059
New York	0.077***	0.060	0.081	0.103
Philadelphia	0.098***	0.101	0.100	0.084
Cleveland	0.089***	0.093	0.092	0.070
Richmond	0.066***	0.086	0.059	0.038
Atlanta	0.087***	0.076	0.105	0.052
Chicago	0.097***	0.093	0.110	0.067
St. Louis	0.081***	0.076	0.087*	0.070
Minneapolis	0.076***	0.080	0.085**	0.038
Kansas City	0.084***	0.085	0.088*	0.070
Dallas	0.082***	0.062	0.097	0.083
San Francisco	0.092***	0.086	0.108	0.057

This table shows the results of the analysis on commercial bank level when constraining individual reweighting to 10% of the original bank size. $r_{\text{MVP}} - r_{\text{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

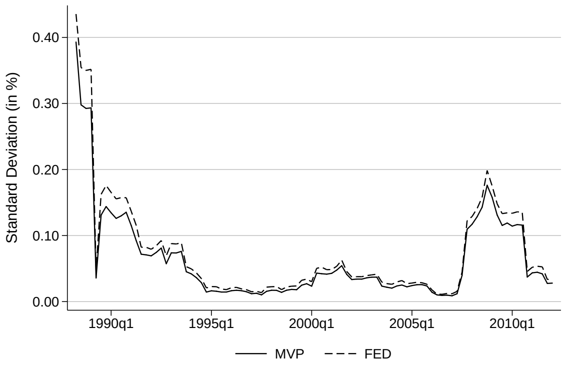
Table 6.C.2: Portfolio optimization using limited reweighting of 10% - Weight

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−0.6***	−0.6	−0.7***	−0.4
New York	−0.4***	−0.6	−0.2	−0.5
Philadelphia	−0.4***	−0.2	−0.4	−0.6
Cleveland	−0.2***	−0.5**	−0.1	−0.0
Richmond	−0.1***	−0.1	−0.1	−0.0
Atlanta	−0.0***	0.2	−0.0	−0.3
Chicago	−0.2***	−0.2	−0.1	−0.5
St. Louis	0.0***	0.3	−0.1	−0.3
Minneapolis	−0.2***	−0.2	−0.2	−0.1
Kansas City	0.1***	0.2*	0.0	0.0
Dallas	−0.2***	−0.3	−0.1	−0.0
San Francisco	−0.2***	−0.2	−0.3	−0.1
MVP 5% - FRD 5%				
Boston	−0.6***	−0.6	−0.6***	−0.4
New York	−0.4***	−0.6	−0.2	−0.5
Philadelphia	−0.4***	−0.2	−0.4	−0.6
Cleveland	−0.2***	−0.5**	−0.1	−0.0
Richmond	−0.1***	−0.1	−0.1	−0.0
Atlanta	0.0***	0.2	−0.0	−0.3
Chicago	−0.2***	−0.2	−0.1	−0.4
St. Louis	0.0***	0.3	−0.1	−0.3
Minneapolis	−0.2***	−0.2	−0.2	−0.1
Kansas City	0.1***	0.2*	0.1	0.0
Dallas	−0.2***	−0.3	−0.1	0.0
San Francisco	−0.2***	−0.2	−0.3	−0.0

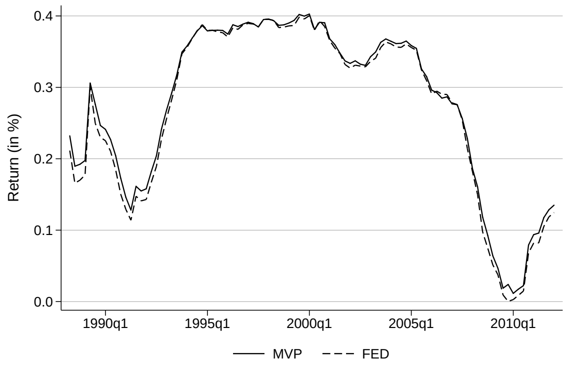
This table shows the results of the analysis on commercial bank level when constraining individual reweighting to 10% of the original bank size. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 6.C.2: Portfolio optimization using limited reweighting - 20%

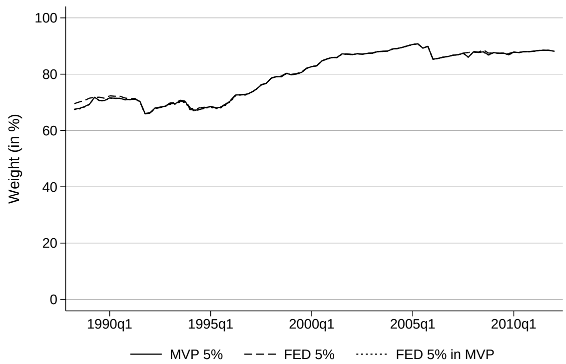
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP when constraining individual reweighting to 20% of the original bank size: panel 6.C.2a and 6.C.2b display the difference in the risk and return of each portfolio. Panel 6.C.2c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.C.3: Portfolio optimization using limited reweighting of 20% - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\text{MVP}} - r_{\text{FRD}}$				
Boston	0.008***	0.017	0.003	−0.001
New York	0.009***	0.012	0.008	0.004
Philadelphia	−0.001***	0.010*	−0.019*	0.035*
Cleveland	0.007***	0.012**	0.004	0.004
Richmond	0.005***	0.010	0.002	0.004
Atlanta	0.005***	0.008	−0.002	0.019
Chicago	0.008***	0.012	0.002	0.018
St. Louis	0.004***	0.005**	0.001	0.012
Minneapolis	−0.009***	0.002	−0.017*	−0.007
Kansas City	0.004***	0.010*	−0.000	0.007
Dallas	0.009***	0.022	0.001	0.005
San Francisco	0.003***	0.014**	−0.006	0.007
$(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$				
Boston	0.142***	0.139	0.152	0.116*
New York	0.154***	0.121	0.161***	0.205*
Philadelphia	0.202***	0.209***	0.207*	0.169
Cleveland	0.175***	0.178	0.184***	0.140
Richmond	0.125***	0.162	0.115	0.074
Atlanta	0.166***	0.153**	0.194**	0.105
Chicago	0.188***	0.188	0.203***	0.140
St. Louis	0.163***	0.150**	0.176***	0.149
Minneapolis	0.156***	0.163	0.175***	0.076
Kansas City	0.176***	0.181*	0.181***	0.148
Dallas	0.167***	0.127*	0.196*	0.164
San Francisco	0.185***	0.180***	0.210*	0.113

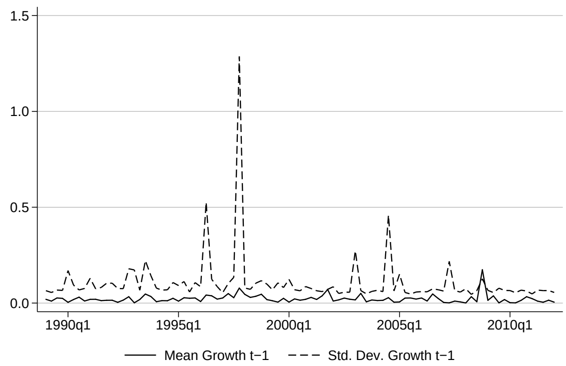
This table shows the results of the analysis on commercial bank level when constraining individual reweighting to 20% of the original bank size. $r_{\text{MVP}} - r_{\text{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Table 6.C.4: Portfolio optimization using limited reweighting of 20% - Weight

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−1.4***	−1.2**	−1.6***	−0.9
New York	−1.0***	−1.4**	−0.6	−1.3
Philadelphia	−1.0***	−0.5	−1.3**	−1.3
Cleveland	−0.5***	−1.1**	−0.2	−0.0
Richmond	−0.1***	−0.2	−0.2	−0.0
Atlanta	0.0***	0.4	−0.0	−0.6
Chicago	−0.3***	−0.4	−0.1	−1.0
St. Louis	−0.1***	0.5	−0.2	−0.7
Minneapolis	−0.4***	−0.4	−0.4	−0.2
Kansas City	0.2***	0.4**	0.2	0.0
Dallas	−0.3***	−0.5	−0.2	−0.1
San Francisco	−0.5***	−0.4*	−0.7	−0.1
MVP 5% - FRD 5%				
Boston	−1.3***	−1.2**	−1.5***	−0.9
New York	−0.9***	−1.4**	−0.6	−1.1
Philadelphia	−0.9***	−0.2	−1.2**	−1.3
Cleveland	−0.5***	−1.0**	−0.2	−0.0
Richmond	−0.1***	−0.1	−0.2	−0.0
Atlanta	0.1***	0.4	−0.0	−0.5
Chicago	−0.3***	−0.4	−0.1	−0.9
St. Louis	0.0***	0.6**	−0.2	−0.6
Minneapolis	−0.4***	−0.4	−0.4	−0.2
Kansas City	0.3***	0.5***	0.2	0.1
Dallas	−0.2***	−0.4	−0.1	−0.1
San Francisco	−0.5***	−0.4*	−0.7	−0.0

This table shows the results of the analysis on commercial bank level when constraining individual reweighting to 20% of the original bank size. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

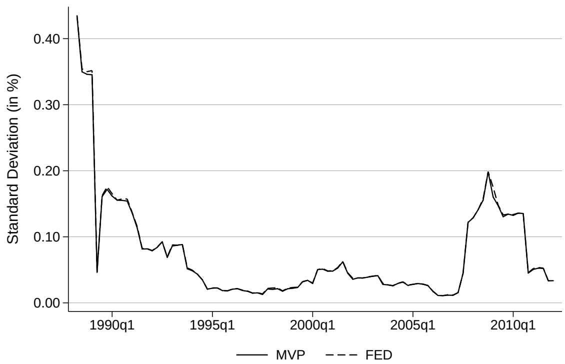
Figure 6.C.3: Limited reweighting - Mean and standard deviation of growth variable



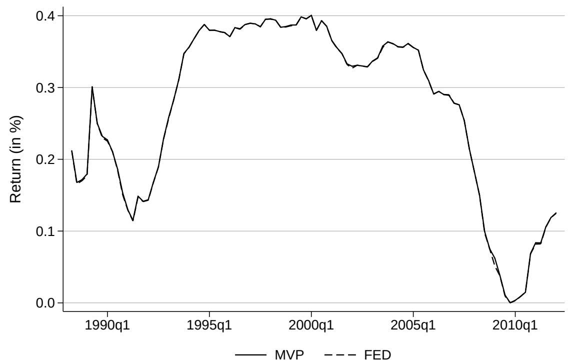
The figure shows the mean and standard deviation of the asset growth in the previous quarter, which are used as a time-varying constraint determining how much the supervisor can alter the assets of individual banks.

Figure 6.C.4: Portfolio optimization using limited reweighting - Mean of growth

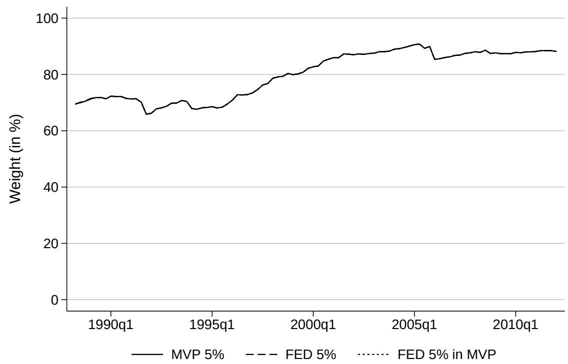
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP when using the mean asset growth as a time-varying constraint for individual reweighting of banks: panel 6.C.4a and 6.C.4b display the difference in the risk and return of each portfolio. Panel 6.C.4c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.C.5: Portfolio optimization using limited reweighting - Mean of growth - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\mathrm{MVP}} - r_{\mathrm{FRD}}$				
Boston	0.003***	0.009	0.000	-0.000
New York	0.002***	0.002	0.001	0.001
Philadelphia	0.002***	0.002	-0.003	0.018
Cleveland	0.001***	0.001	0.000	0.000
Richmond	0.001***	0.001	0.000	0.000
Atlanta	0.001***	0.001	-0.000	0.006
Chicago	0.001***	0.001	0.000	0.003
St. Louis	0.000***	0.000	0.000	0.001
Minneapolis	-0.003***	-0.000	-0.002	-0.012
Kansas City	0.000***	0.001	-0.000	0.001
Dallas	0.002***	0.005	0.000	0.001
San Francisco	0.001***	0.003	-0.002	0.002
$(\sigma_{\mathrm{FRD}} - \sigma_{\mathrm{MVP}})/\sigma_{\mathrm{FRD}}$				
Boston	0.041***	0.076	0.024	0.013
New York	0.031***	0.024	0.033	0.037
Philadelphia	0.040***	0.043	0.039	0.038
Cleveland	0.019***	0.022	0.018	0.017
Richmond	0.021***	0.027	0.021	0.010
Atlanta	0.028***	0.020	0.035	0.024
Chicago	0.019***	0.015	0.023	0.015
St. Louis	0.013***	0.011	0.015	0.012
Minneapolis	0.027***	0.013	0.024	0.069
Kansas City	0.011***	0.011	0.012	0.011
Dallas	0.018***	0.017	0.017	0.022
San Francisco	0.046***	0.038	0.059	0.024

This table shows the results of the analysis on commercial bank level when using the mean asset growth as a time-varying constraint for individual reweighting of banks. $r_{\mathrm{MVP}} - r_{\mathrm{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\mathrm{FRD}} - \sigma_{\mathrm{MVP}})/\sigma_{\mathrm{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

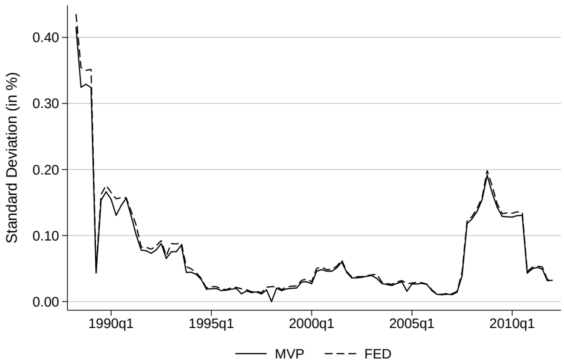
Table 6.C.6: Portfolio optimization using limited reweighting - Mean of growth - Weight

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−1.0***	−2.4	−0.2	−0.1
New York	−0.2***	−0.2	−0.1	−0.2
Philadelphia	−0.2***	−0.1	−0.2	−0.2
Cleveland	−0.0***	−0.1	−0.0	−0.0
Richmond	−0.0***	−0.0	−0.0	0.0
Atlanta	−0.0***	0.0	−0.0	−0.1
Chicago	−0.0***	−0.0	−0.0	−0.1
St. Louis	0.0***	0.1	−0.0	−0.0
Minneapolis	−0.8***	−0.0	−0.0	−5.2
Kansas City	0.0***	0.0	0.0	−0.0
Dallas	−0.0***	−0.1	−0.0	−0.0
San Francisco	−0.1***	−0.0	−0.2	−0.0
MVP 5% - FRD 5%				
Boston	−0.7***	−1.6	−0.2	−0.1
New York	−0.1***	−0.2	−0.1	−0.1
Philadelphia	−0.1***	−0.1	−0.2	−0.1
Cleveland	−0.0***	−0.1	−0.0	−0.0
Richmond	−0.0***	−0.0	−0.0	0.0
Atlanta	−0.0***	0.0	−0.0	−0.1
Chicago	−0.0***	−0.0	−0.0	−0.1
St. Louis	0.0***	0.1	−0.0	−0.0
Minneapolis	−0.7***	−0.0	−0.0	−4.4
Kansas City	0.0***	0.0	0.0	−0.0
Dallas	−0.0***	−0.1	−0.0	−0.0
San Francisco	−0.1***	−0.0	−0.2	−0.0

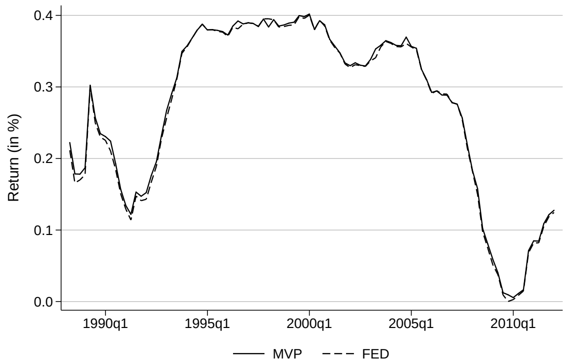
This table shows the results of the analysis on commercial bank level when using the mean asset growth as a time-varying constraint for individual reweighting of banks. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 6.C.5: Portfolio optimization using limited reweighting - Std. dev. of growth

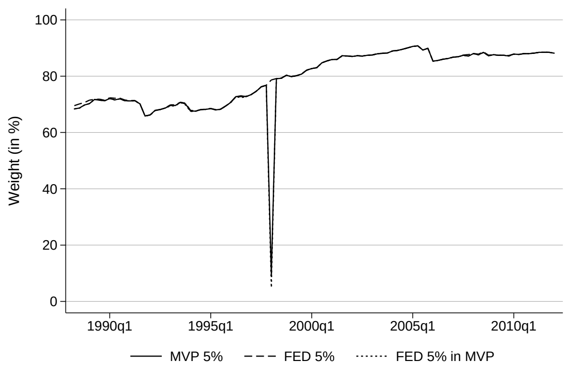
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?



(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP level when using the standard deviation of asset growth as a time-varying constraint for individual reweighting of banks: panel 6.C.5a and 6.C.5b display the difference in the risk and return of each portfolio. Panel 6.C.5c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.C.7: Portfolio optimization using limited reweighting - Std. dev. of growth - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\mathrm{MVP}} - r_{\mathrm{FRD}}$				
Boston	0.008***	0.021	0.002	-0.001
New York	0.008***	0.011	0.008	0.002
Philadelphia	0.005***	0.015*	-0.024	0.079
Cleveland	0.005***	0.010	0.002	0.001
Richmond	0.004***	0.010	0.001	0.002
Atlanta	0.005***	0.008	-0.002	0.022
Chicago	0.005***	0.008	0.001	0.010
St. Louis	0.001***	0.000	0.001	0.003
Minneapolis	-0.009***	-0.000	-0.014	-0.015
Kansas City	0.002***	0.006	-0.000	0.003
Dallas	0.010***	0.028	0.001	0.002
San Francisco	-0.002***	0.008	-0.012	0.009
$(\sigma_{\mathrm{FRD}} - \sigma_{\mathrm{MVP}})/\sigma_{\mathrm{FRD}}$				
Boston	0.126***	0.169	0.110	0.080
New York	0.142***	0.110	0.163*	0.147
Philadelphia	0.215***	0.267***	0.178	0.215
Cleveland	0.112***	0.157	0.093	0.065
Richmond	0.094***	0.138	0.080	0.040
Atlanta	0.154***	0.134***	0.180**	0.119
Chicago	0.128***	0.113	0.156**	0.070
St. Louis	0.092***	0.092	0.104**	0.053
Minneapolis	0.120***	0.088	0.153***	0.087
Kansas City	0.109***	0.109	0.124*	0.061
Dallas	0.095***	0.115	0.088	0.074
San Francisco	0.223***	0.230**	0.246**	0.134

This table shows the results of the analysis on commercial bank level when using the standard deviation of asset growth as a time-varying constraint for individual reweighting of banks. $r_{\mathrm{MVP}} - r_{\mathrm{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\mathrm{FRD}} - \sigma_{\mathrm{MVP}})/\sigma_{\mathrm{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

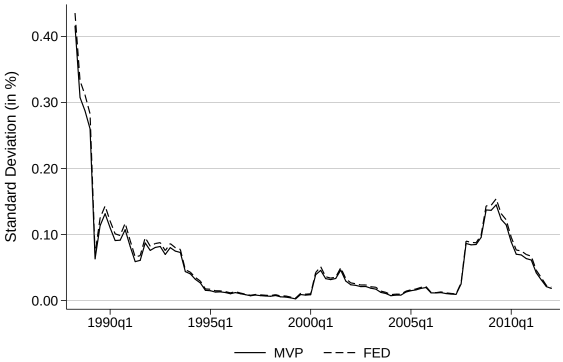
Table 6.C.8: Portfolio optimization using limited reweighting - Std. dev. of growth - Weight

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−2.3***	−4.5**	−1.2***	−0.6
New York	−0.9***	−1.3**	−0.7	−0.8
Philadelphia	−4.1***	−4.1	−3.2**	−6.6
Cleveland	−0.9***	−2.5**	−0.1	−0.0
Richmond	−0.7***	−2.0	−0.1	−0.0
Atlanta	−1.2***	0.3	−1.3	−4.4
Chicago	−0.8***	−0.2	−1.3	−0.5
St. Louis	−0.4***	−0.9	−0.1	−0.2
Minneapolis	−2.2***	−0.2	−2.6	−5.3
Kansas City	0.1***	0.2	0.1	0.0
Dallas	−0.7***	−1.8	−0.1	−0.0
San Francisco	−5.4***	−4.6*	−6.0	−5.2
MVP 5% - FRD 5%				
Boston	−1.9***	−3.4**	−1.2***	−0.6
New York	−0.8***	−1.3**	−0.5	−0.7
Philadelphia	−3.2***	−2.2	−3.0**	−6.0
Cleveland	−0.8***	−2.2**	−0.1	−0.0
Richmond	−0.6***	−1.6	−0.1	−0.0
Atlanta	−1.0***	0.4	−1.3	−3.2
Chicago	−0.7***	−0.2	−1.2	−0.4
St. Louis	−0.3***	−0.6	−0.1	−0.2
Minneapolis	−1.9***	−0.2	−2.4	−4.5
Kansas City	0.1***	0.2*	0.1	0.0
Dallas	−0.6***	−1.5	−0.1	0.0
San Francisco	−4.9***	−4.0*	−5.7	−3.9

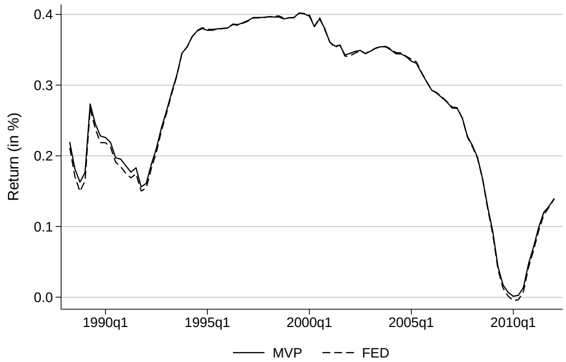
This table shows the results of the analysis on commercial bank level when using the standard deviation of asset growth as a time-varying constraint for individual reweighting of banks. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure 6.C.6: Portfolio optimization using dynamic limited reweighting of 10%

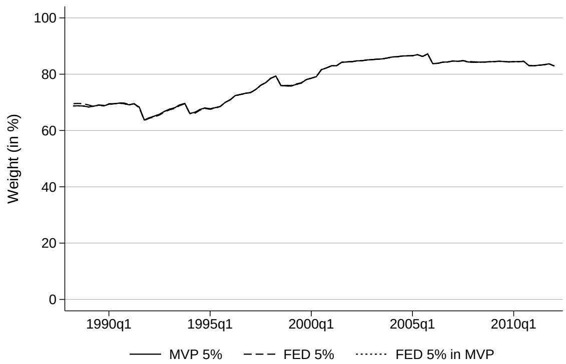
(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?

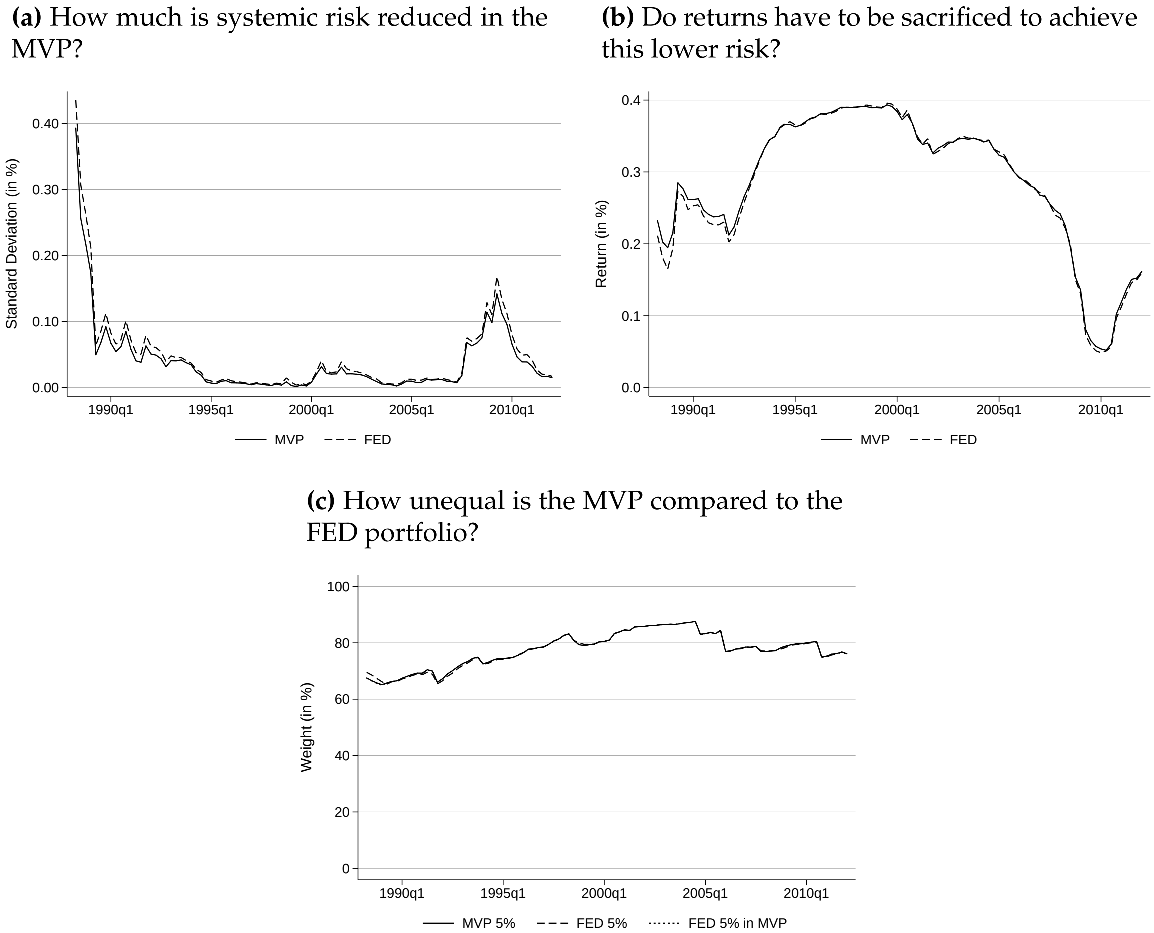


(c) How unequal is the MVP compared to the FED portfolio?



The figure shows the comparison between the FED portfolio and the hypothetical MVP when constraining individual reweighting to 10% of the bank size in the previous quarter's MVP: panel 6.C.6a and 6.C.6b display the difference in the risk and return of each portfolio. Panel 6.C.6c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Figure 6.C.7: Portfolio optimization using dynamic limited reweighting of 20%



The figure shows the comparison between the FED portfolio and the hypothetical MVP when constraining individual reweighting to 20% of the bank size in the previous quarter's MVP: panel 6.C.7a and 6.C.7b display the difference in the risk and return of each portfolio. Panel 6.C.7c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

Table 6.C.9: Portfolio optimization keeping the largest 5% at their actual weight - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\text{MVP}} - r_{\text{FRD}}$				
Boston	0.060***	0.094***	0.046***	0.025
New York	−0.001***	0.041	−0.009*	−0.076*
Philadelphia	0.005***	0.063***	−0.032*	−0.010
Cleveland	0.005***	−0.011	0.015**	0.008
Richmond	−0.028***	−0.032	−0.026***	−0.028
Atlanta	−0.014***	−0.044	−0.016*	0.061**
Chicago	−0.029***	−0.061***	−0.017	0.006
St. Louis	−0.035***	−0.036***	−0.033***	−0.038
Minneapolis	−0.037***	−0.114***	−0.002*	0.028
Kansas City	−0.052***	−0.072***	−0.035***	−0.059**
Dallas	−0.010***	0.047**	−0.035***	−0.062**
San Francisco	−0.018***	−0.090***	0.027*	0.001*
$(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$				
Boston	0.634***	0.784***	0.608***	0.373*
New York	0.623***	0.646***	0.620***	0.579***
Philadelphia	0.798***	0.975***	0.869***	0.157
Cleveland	0.738***	0.966***	0.747***	0.183
Richmond	0.651***	0.994***	0.476***	0.425
Atlanta	0.927***	1.000***	0.999***	0.520*
Chicago	0.905***	0.950***	0.893***	0.844***
St. Louis	0.998***	1.000***	1.000***	0.987***
Minneapolis	0.876***	0.990***	0.943***	0.391***
Kansas City	1.000***	1.000***	1.000***	1.000***
Dallas	0.965***	0.908***	1.000***	0.980***
San Francisco	0.898***	0.933***	0.967***	0.594***

This table shows the results of the analysis on commercial bank level when constraining the largest 5% of banks to their individual current weigh, while allowing unlimited reweighting for the remaining banks. $r_{\text{MVP}} - r_{\text{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Table 6.C.10: Portfolio optimization keeping the largest 5% at their actual weight - Weight

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	0.0***	0.0	0.0	0.0
New York	0.0***	0.0	0.0	0.0
Philadelphia	0.0***	0.0	0.0	0.0
Cleveland	0.0***	0.0	0.0	0.0
Richmond	0.0***	0.0	0.0	0.0
Atlanta	0.0***	0.0	0.0	0.0
Chicago	0.0***	0.0	0.0	0.0
St. Louis	0.0***	0.0	0.0	0.0
Minneapolis	0.0***	0.0	0.0	0.0
Kansas City	0.0***	0.0	0.0	0.0
Dallas	0.0***	0.0	0.0	0.0
San Francisco	0.0***	0.0	0.0	0.0
MVP 5% - FRD 5%				
Boston	7.8***	13.3***	5.7***	1.6*
New York	8.3***	10.4***	5.3***	13.2***
Philadelphia	13.8***	16.3***	14.0***	7.2***
Cleveland	7.9***	13.1***	6.1***	1.5***
Richmond	7.0***	7.5***	7.6***	4.1***
Atlanta	6.9***	5.8***	3.7***	19.9***
Chicago	9.8***	13.4***	4.9***	17.6***
St. Louis	8.6***	5.1***	7.4***	20.5***
Minneapolis	12.4***	18.0***	10.0***	7.3***
Kansas City	5.6***	3.9***	5.1***	11.1***
Dallas	11.6***	17.9***	5.3***	17.7***
San Francisco	7.3***	8.4***	6.6***	7.2***

This table shows the results of the analysis on commercial bank level when constraining the largest 5% of banks to their individual current weigh, while allowing unlimited reweighting for the remaining banks. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6.C.11: Portfolio optimization keeping the largest 5% between 50% and 60% - Risk-return trade-off

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
$r_{\text{MVP}} - r_{\text{FRD}}$				
Boston	0.060***	0.140***	0.003	0.058
New York	0.051***	0.128***	0.021***	−0.028*
Philadelphia	−0.049***	0.057***	−0.178***	0.123**
Cleveland	0.032***	0.067***	−0.011***	0.093***
Richmond	−0.006***	0.034***	−0.026***	−0.037
Atlanta	0.006***	0.001	−0.017***	0.091***
Chicago	0.044***	0.085***	0.005	0.076***
St. Louis	0.002***	0.010***	−0.022***	0.059***
Minneapolis	−0.074***	−0.033**	−0.108***	−0.060
Kansas City	−0.017***	−0.029**	−0.013**	−0.003
Dallas	0.031***	0.102**	−0.013***	0.005
San Francisco	−0.013***	0.015*	−0.035***	−0.004*
$(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$				
Boston	0.994***	1.000***	1.000***	0.964***
New York	1.000***	1.000***	1.000***	1.000***
Philadelphia	1.000***	1.000***	1.000***	1.000***
Cleveland	1.000***	1.000***	1.000***	1.000***
Richmond	1.000***	1.000***	1.000***	1.000***
Atlanta	1.000***	1.000***	1.000***	1.000***
Chicago	1.000***	1.000***	1.000***	1.000***
St. Louis	1.000***	1.000***	1.000***	1.000***
Minneapolis	1.000***	1.000***	1.000***	1.000***
Kansas City	1.000***	1.000***	1.000***	1.000***
Dallas	1.000***	1.000***	1.000***	1.000***
San Francisco	1.000***	1.000***	1.000***	1.000***

This table shows the results of the analysis on commercial bank level when constraining the largest 5% of banks to have a weight between 50% and 60%. $r_{\text{MVP}} - r_{\text{FRD}}$ measures the difference between returns of the actual portfolio and the returns of the minimum variance portfolio. $(\sigma_{\text{FRD}} - \sigma_{\text{MVP}})/\sigma_{\text{FRD}}$ measures the relative difference in portfolio risk between the actual and minimum variance portfolio. A score of 1 indicates that the risk has been effectively eliminated, while a score of 0 indicates that no improvements are possible. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** p<0.01, ** p<0.05, * p<0.1.

Table 6.C.12: Portfolio optimization keeping the largest 5% between 50% and 60% - Weight

	1984Q4 - 2010Q4	1984Q4 - 1993Q4	1994Q1 - 2006Q4	2007Q1 - 2010Q4
FRD 5% in MVP - FRD 5%				
Boston	−24.2***	−14.9***	−29.0***	−29.9***
New York	−30.6***	−29.4***	−35.2***	−18.3***
Philadelphia	−13.1***	0.1***	−15.9***	−34.1***
Cleveland	−30.6***	−17.5***	−34.8***	−46.8***
Richmond	−33.0***	−22.1***	−37.3***	−44.0***
Atlanta	−19.3***	−14.5***	−22.1***	−21.1***
Chicago	−17.3***	−10.0***	−22.8***	−15.9***
St. Louis	−3.9***	1.4***	−8.4***	−1.8***
Minneapolis	−18.1***	−5.7***	−20.0***	−40.3***
Kansas City	1.2***	6.5***	−2.3***	0.2***
Dallas	−7.1***	−5.9***	−8.9***	−3.9***
San Francisco	−26.1***	−30.0***	−23.0***	−27.2***
MVP 5% - FRD 5%				
Boston	−16.9***	−8.3***	−24.5***	−12.1***
New York	−27.1***	−24.8***	−34.1***	−9.5***
Philadelphia	−9.8***	2.2***	−14.7***	−21.2***
Cleveland	−29.5***	−16.6***	−34.6***	−43.2***
Richmond	−30.7***	−19.0***	−36.1***	−40.0***
Atlanta	−17.5***	−12.2***	−21.9***	−15.7***
Chicago	−16.6***	−9.6***	−22.8***	−12.8***
St. Louis	−2.8***	2.4***	−7.7***	0.9
Minneapolis	−17.1***	−4.8***	−19.5***	−37.7***
Kansas City	2.0***	7.1***	−1.7***	2.2***
Dallas	−5.9***	−4.1***	−8.3***	−1.9
San Francisco	−25.0***	−28.4***	−22.7***	−24.6***

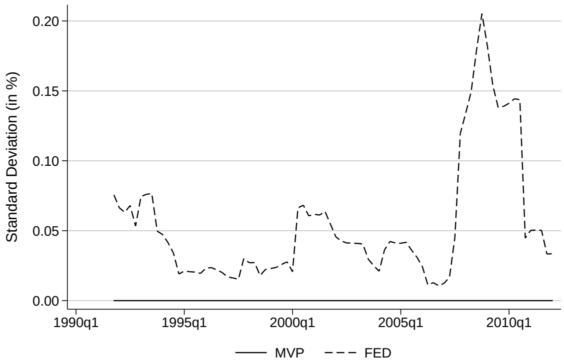
This table shows the results of the analysis on commercial bank level when constraining the largest 5% of banks to have a weight between 50% and 60%. *FRD 5% in MVP - FRD 5%* measures the average difference between weight of the top-5% in the actual portfolio with the weight of the actual top-5% in the minimum variance portfolio. *MVP 5% - FRD 5%* measures the average difference between the weight of the top-5% in the actual portfolio with the weight of the new top-5% in the minimum variance portfolio. Kolmogorov-Smirnov tests are performed to see if the distributions of the actual and minimum variance portfolio are different from each other. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

6.C.2 Portfolio Optimization using Listed BHCs

Lastly, we check whether the results hold if we only consider BHCs that had publicly traded equity, using the database of the Federal Reserve Bank of New York (2013). Banks without a listing are removed and only those that do have one are considered for the portfolio the investor can manage. Figure 6.C.8 shows the results, which are again similar to the baseline scenario.

Figure 6.C.8: Portfolio optimization using listed BHCs

(a) How much is systemic risk reduced in the MVP?



(b) Do returns have to be sacrificed to achieve this lower risk?

(c) How unequal is the MVP compared to the FED portfolio?

The figure shows the comparison between the FED portfolio and the hypothetical MVP when using only BHCs that have publicly traded equity: panel 6.C.8a and 6.C.8b display the difference in the risk and return of each portfolio. Panel 6.C.8c shows how the weights are distributed in each portfolio by plotting their concentration ratios, as well as the weights that the current largest banks have in the MVP.

References

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review 93*(1), pp. 113–132.

Abadie, A. and J. Gardeazabal (2008). Terrorism and the world economy. *European Economic Review 52*, pp. 1–27.

Abowd, J. M., R. H. Creecy, and F. Kramarz (2002). Computing person and firm effects using linked longitudinal employer-employee data. Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

Acharya, V. V. (2009). A theory of systemic risk and design of prudential bank regulation. *Journal of Financial Stability 5*(3), pp. 224–255.

Acharya, V. V., D. Anginer, and A. J. Warburton (2014). The end of market discipline? Investor expectations of implicit government guarantees. Working Paper. <http://ssrn.com/abstract=1961656>.

Acharya, V. V., R. F. Engle, and M. P. Richardson (2012). Capital Shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review 102*(3), pp. 59–64.

Acharya, V. V., L. H. Pedersen, T. Philippon, and M. P. Richardson (2012). Measuring systemic risk. CEPR Discussion Papers 8824, C.E.P.R. Discussion Papers.

Acharya, V. V. and T. Yorulmazer (2007). Too many to fail – An analysis of time-inconsistency in bank closure policies. *Journal of Financial Intermediation 16*(1), pp. 1–31.

Adams, R. M., K. P. Brevoort, and E. K. Kiser (2007). Who competes with whom? The case of depository institutions. *Journal of Industrial Economics 55*(1), pp. 141–167.

Adrian, T. and M. K. Brunnermeier (2011). CoVaR. Working Paper 17454, National Bureau of Economic Research.

Aitken, B. J. and A. E. Harrison (1999). Do domestic firms benefit from direct foreign investment? Evidence from Venezuela. *American Economic Review 89*(3), pp. 605–618.

Alegria, C. and K. Schaeck (2009). On measuring concentration in banking systems. *Finance Research Letters 5*, pp. 59–67.

Allen, F., A. Babus, and E. Carletti (2012). Asset commonality, debt maturity and systemic risk. *Journal of Financial Economics 104*(3), pp. 519–534.

Allen, F. and E. Carletti (2008). Mark-to-market accounting and liquidity pricing. *Journal of Accounting and Economics 45*(2-3), pp. 358–378.

Ashcraft, A. B. (2008). Does the market discipline banks? New evidence from regulatory capital mix. *Journal of Financial Intermediation 17*(4), pp. 543–561.

Avery, R. B., T. M. Belton, and M. A. Goldberg (1988). Market discipline in regulating bank risk: New evidence from the capital markets. *Journal of Money, Credit and Banking 20*(4), pp. 597–610.

Baele, L., V. De Bruyckere, O. De Jonghe, and R. Vander Vennet (2014). Do stock markets discipline US Bank Holding Companies: Just monitoring, or also influencing? *North American Journal of Economics and Finance 29*, pp. 124–145.

Bekaert, G., M. Ehrmann, M. Fratzscher, and A. J. Mehl (2011). Global crises and equity market contagion. NBER Working Paper 17121.

Berger, A. N. and C. H. Bouwman (2013). How does capital affect bank performance during financial crises? *Journal of Financial Economics 109*(1), pp. 146–176.

Berger, A. N., R. S. Demsetz, and P. E. Strahan (1999). The consolidation of the financial services industry: Causes, consequences, and implications for the future. *Journal of Banking & Finance 23*(2-4), pp. 135–194.

Berger, A. N. and R. Turk-Ariss (2014). Do depositors discipline banks and did government actions during the recent crisis reduce this discipline? An international perspective. *Journal of Financial Services Research*, forthcoming.

Berrebi, C. and E. Klor (2010). The impact of terrorism on the defence industry. *Economica 77*, pp. 518–543.

Bertay, A. C., A. Demirgüç-Kunt, and H. Huizinga (2013). Do we need big banks? Evidence on performance, strategy and market discipline. *Journal of Financial Intermediation 22*(4), pp. 532–558.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), pp. 249–275.

Billett, M. T., J. A. Garfinkel, and E. S. O'Neal (1998). The cost of market versus regulatory discipline in banking. *Journal of Financial Economics 48*(3), pp. 333–358.

Bisias, D., M. Flood, A. W. Lo, and S. Valavanis (2012). A survey of systemic risk analytics. *Annual Review of Financial Economics 4*(1), pp. 255–296.

Blank, S., C. M. Buch, and K. Neugebauer (2009). Shocks at large banks and banking sector distress: The banking granular residual. *Journal of Financial Stability 5*(4), pp. 353–373.

Blomberg, S., G. D. Hess, and A. Orphanides (2004). The macroeconomic consequences of terrorism. *Journal of Monetary Economics 51*, pp. 1007–1032.

Blomberg, S. B. and G. D. Hess (2009). Estimating the macroeconomic consequence of 9/11. *Peace Economics, Peace Science, and Public Policy 15*(2), pp. 1–26.

Boot, A. and L. Ratnovski (2012). Banking and trading. IMF Working Paper 12/238.

Bottom, W. P., K. Ladha, and G. J. Miller (2002). Propagation of individual bias through group judgment: Error in the treatment of asymmetrically informative signals. *Journal of Risk and Uncertainty 25*(2), pp. 147–63.

Bozzoli, C. and C. Müller (2011). Perceptions and attitudes following a terrorist shock: Evidence from the UK. *European Journal of Political Economy 27*, pp. 89–106.

Brewer III, E. and T. H. Mondschean (1994). An empirical test of the incentive effects of deposit insurance: The case of junk bonds at savings and loan associations. *Journal of Money, Credit and Banking 26*(1), pp. 146–64.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences 106*(30), pp. 12267–12272.

Brown, C. O. and I. S. Dinç (2011). Too many to fail? Evidence of regulatory forbearance when the banking sector is weak. *Review of Financial Studies 24*(4), pp. 1378–1405.

Brown, J. R., J. D. Cummins, C. M. Lewis, and R. Wei (2004). An empirical analysis of the economic impact of federal terrorism reinsurance. *Journal of Monetary Economics 51*(5), pp. 861–898.

Brownlees, C. T. and R. F. Engle (2012). Volatility, correlation and tails for systemic risk measurement. Working paper. <http://ssrn.com/abstract=1611229>.

Brunnermeier, M. K., G. Dong, and D. Palia (2012). Banks' non-interest income and systemic risk. Working paper. <http://ssrn.com/abstract=1786738>.

Carter, D. A. and B. J. Simkins (2004). The market's reaction to unexpected, catastrophic events: The case of airline stock returns and the September 11th attacks. *Quarterly Review of Economics and Finance 44*, pp. 539–558.

Carvalho, V. and X. Gabaix (2013). The great diversification and its undoing. *American Economic Review 103*(5), pp. 1697–1727.

Chen, A. H. and T. F. Siems (2004). The effects of terrorism on global capital markets. *European Journal of Political Economy 20*, pp. 349–366.

Chen, S.-S., K. W. Ho, C.-f. Lee, and G. H. H. Yeo (2000). Investment opportunities, free cash flow and market reaction to international joint ventures. *Journal of Banking & Finance 24*(11), pp. 1747 – 1765.

Chesney, M., G. Reshetarb, and M. Karaman (2011). The impact of terrorism on financial markets: An empirical study. *Journal of Banking & Finance 35*(2), pp. 253–267.

Claeys, S. and K. Schoors (2007). Bank supervision Russian style: Evidence of conflicts between micro- and macro-prudential concerns. *Journal of Comparative Economics 35*(3), pp. 630–657.

Colin Cameron, A. and D. L. Miller (2014). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, forthcoming.

Collender, R. N. and S. Shaffer (2003). Local bank office ownership, deposit control, market structure, and economic growth. *Journal of Banking & Finance 27*(1), pp. 27–57.

Cook, D. O. and L. J. Spellman (1994). Repudiation risk and restitution costs: Toward understanding premiums on insured deposits. *Journal of Money, Credit and Banking 26*(3), pp. 439–59.

Correa, R., H. Sapriza, and A. Zlate (2012). Liquidity shocks, dollar funding costs, and the bank lending channel during the European sovereign crisis. International Finance Discussion Papers 1059, Board of Governors of the Federal Reserve System.

Craig, B. R. and V. Dinger (2013). Deposit market competition, wholesale funding, and bank risk. *Journal of Banking & Finance 37*(9), pp. 3605–3622.

Craig MacKinlay, A. (1997). Event studies in economics and finance. *Journal of Economic Literature 35*, pp. 13–39.

Cubillas, E., A. R. Fonseca, and F. González (2012). Banking crises and market discipline: International evidence. *Journal of Banking & Finance 36*(8), pp. 2285–2298.

De Bondt, W. F. and R. Thaler (1985). Does the stock market overreact? *Journal of Finance 40*(3), pp. 793–805.

De Bondt, W. F. and R. H. Thaler (1990). Do security analysts overreact? *American Economic Review 80*(2), pp. 52–57.

De Jonghe, O. (2010). Back to the basics in banking? A micro-analysis of banking system stability. *Journal of Financial Intermediation 19*(3), pp. 387–417.

Degryse, H. and G. Nguyen (2007). Interbank exposures: An empirical examination of contagion risk in the Belgian banking system. *International Journal of Central Banking 3*(2), pp. 123–171.

DeMiguel, V., L. Garlappi, and R. Uppal (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *Review of Financial Studies 22*(5), pp. 1915–1953.

Demirgüç-Kunt, A. and H. Huizinga (2013). Are banks too big to fail or too big to save? International evidence from equity prices and CDS spreads. *Journal of Banking & Finance 37*(3), pp. 875–894.

Demirgüç-Kunt, A. and H. Huizinga (2004). Market discipline and deposit insurance. *Journal of Monetary Economics 51*(2), pp. 375–399.

DeYoung, R. (2005). The performance of Internet-based business models:Evidence from the banking industry. *Journal of Business 78*(3), pp. 893–948.

DeYoung, R. and I. Hasan (1998). The performance of de novo commercial banks: A profit efficiency approach. *Journal of Banking & Finance 22*, pp. 565–587.

DeYoung, R. and W. C. Hunter (2002). Deregulation, the Internet, and the competitive viability of large banks and community banks. In B. Gup (Ed.), *The Future of Banking*. Greenwood Press.

DeYoung, R., W. W. Lang, and D. L. Nolle (2007). How the Internet affects output and performance at community banks. *Journal of Banking & Finance 31*(4), pp. 1033–1060.

DeYoung, R. and K. P. Roland (2001). Product mix and earnings volatility at commercial banks: Evidence from a degree of total leverage model. *Journal of Financial Intermediation 10*(1), 54–84.

DeYoung, R. and G. Torna (2013). Nontraditional banking activities and bank failures during the financial crisis. *Journal of Financial Intermediation 22*(3), pp. 397–421.

Doukas, J. and N. G. Travlos (1988). The effect of corporate multinationalism on shareholders' wealth: Evidence from international acquisitions. *Journal of Finance 43*(5), pp. 1161–1175.

Dowie, J. (1976). On the efficiency and equity of betting markets. *Economica 43*(170), pp. 139–150.

Drakos, K. (2004). Terrorism-induced structural shifts in financial risk: Airline stocks in the aftermath of the September 11th terror attacks. *European Journal of Political Economy 20*, pp. 435–446.

Drakos, K. (2010a). The determinants of terrorist shocks' cross-market transmission. *Journal of Risk Finance 11*(2), pp. 147–163.

Drakos, K. (2010b). Terrorism activity, investor sentiment, and stock returns. *Review of Financial Economics 19*(3), pp. 128–135.

Drakos, K. and C. Müller (2014). On the determinants of terrorism risk concern in europe. *Defence and Peace Economics 25*(3), pp. 291–310.

Dube, A., E. Kaplan, and S. Naidu (2011). Coups, corporations and classified information. *Quarterly Journal of Economics 126*, pp. 1375–1409.

Dungey, M., R. Fry, B. Gonzalez-Hermosillo, and V. Martin (2005). Empirical modelling of contagion: a review of methodologies. *Quantitative Finance 5*(1), pp. 9–24.

Eckstein, Z. and D. Tsiddon (2004). Macroeconomics consequences of terror: Theory and the case of Israel. *Journal of Monetary Economics 51*, pp. 971–1002.

Eldor, R. and R. Melnick (2004). Financial markets and terrorism. *European Journal of Political Economy 20*, pp. 367–386.

Ellis, D. M. and M. J. Flannery (1992). Does the debt market assess large banks' risk? Time series evidence from money center CDs. *Journal of Monetary Economics 30*(3), pp. 481–502.

EM-DAT. The OFDA/CRED International Disaster Database. Université Catholique de Louvain, Brussels (Belgium). www.cred.be/emdat.

Enders, W., A. Sachsida, and T. Sandler (2006). The impact of transnational terrorism on U.S. foreign direct investment. *Political Research Quarterly 59*(4), pp. 517–531.

Engle, R. F., E. Jondeau, and M. Rockinger (2014). Systemic risk in Europe. *Review of Finance*, forthcoming.

Fahlenbrach, R., R. Prilmeier, and R. M. Stulz (2012). This time is the same: Using bank performance in 1998 to explain bank performance during the recent financial crisis. *Journal of Finance 67*(6), pp. 2139–2185.

Fama, E. F. (1965). The behavior of stock-market prices. *Journal of Business 38*(1), pp. 34–105.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance 25*(2), pp. 383–417.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), pp. 427–465.

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

FDIC (2012). Quarterly Banking Profile First Quarter 2012. Retrieved on July 18, 2014 from: <http://www2.fdic.gov/qbp/2012mar/qbp.pdf>.

FDIC (2014). Summary of Deposits Reporting Instructions. Retrieved on July 18, 2014 from: <http://www2.fdic.gov/sod/pdf/SOD_Instructions.pdf>.

Federal Reserve Bank of New York (2013). CRSP-FRB Link.

Flannery, M. J. (1994). Debt maturity and the deadweight cost of leverage: Optimally financing banking firms. *American Economic Review 84*(1), pp. 320–31.

Flannery, M. J. and S. M. Sorescu (1996). Evidence of bank market discipline in subordinated debenture yields: 1983–1991. *Journal of Finance 51*(4), pp. 1347–1377.

Frey, B. S. and S. Luechinger (2005). Measuring terrorism. In A. Marciano and J.-M. Josselin (Eds.), *Law and the State: A Political Economy Approach: New Horizons in Law and Economics*, pp. 142–181. Cheltenham, UK: Edward Elgar Publishing.

Furfine, C. H. (1999). The microstructure of the Federal Funds market. *Financial Markets, Institutions & Instruments 8*(5), pp. 24–44.

Furfine, C. H. (2003). Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking 35*(1), pp. 111–128.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), pp. 733–772.

Gassebner, M., R. Jong-A-Pin, and J. O. Mierau (2011). Terrorism and cabinet duration. *International Economic Review 52*(4), pp. 1253–1270.

Gennaioli, N. and A. Shleifer (2010). What comes to mind. *Quarterly Journal of Economics 125*(4), pp. 1399–1433.

Greene, W. (2004). Fixed effects and bias due to the incidental parameters problem in the tobit model. *Econometric Reviews 23*(2), pp. 125–147.

Gropp, R., H. Hakenes, and I. Schnabel (2011). Competition, risk-shifting, and public bail-out policies. *Review of Financial Studies 24*(6), pp. 2084–2120.

GTD. The Global Terrorism Database. National Consortium for the Study of Terrorism and Responses to Terrorism, University of Maryland. www.start.umd.edu/gtd/.

Guimarães, P. and P. Portugal (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal 10*(4), pp. 628–649.

Hannan, T. H. and G. A. Hanweck (1988). Bank insolvency risk and the market for large certificates of deposit. *Journal of Money, Credit and Banking 20*(2), pp. 203–11.

Harris, R. S. and D. Ravenscraft (1991). The role of acquisitions in foreign direct investment: Evidence from the U.S. stock market. *Journal of Finance 46*(3), 825–844.

Hasan, I., K. Jackowicz, O. Kowalewski, and Ł. Kozłowski (2013). Market discipline during crisis: Evidence from bank depositors in transition countries. *Journal of Banking & Finance 37*(12), pp. 5436–5451.

Hausman, J. A. and G. K. Leonard (1997). Superstars in the National Basketball Association: Economic value and policy. *Journal of Labor Economics 15*(4), pp. 586–624.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies 65*(2), pp. 261–94.

Heitfield, E. A. (1999). What do interest rate data say about the geography of retail banking markets? *Antitrust Bulletin 44*(2), pp. 333–347.

Helpman, E. (2006). Trade, FDI, and the organization of firms. *Journal of Economic Literature 44*(3), pp. 589–630.

Hernando, I. and M. J. Nieto (2007). Is the Internet delivery channel changing banks' performance? The case of Spanish banks. *Journal of Banking & Finance 31*(4), pp. 1083–1099.

Hirshleifer, D. and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*, pp. 337–386.

Huang, R. (2008). Evaluating the real effect of bank branching deregulation: Comparing contiguous counties across US state borders. *Journal of Financial Economics 87*(3), pp. 678–705.

Huang, R. and L. Ratnovski (2011). The dark side of bank wholesale funding. *Journal of Financial Intermediation 20*(2), pp. 248–263.

Hughes, J. P. and L. J. Mester (2013). Who said large banks don't experience scale economies? Evidence from a risk-return-driven cost function. *Journal of Financial Intermediation 22*(4), pp. 559–585.

IAGS. Institute for the Analysis of Global Security. Retrieved on January 30, 2014 from: <http://www.iags.org/costof911.html>.

Ibragimov, R., D. Jaffee, and J. Walden (2011). Diversification disasters. *Journal of Financial Economics 99*(2), pp. 333–348.

Ioannidou, V. and J. de Dreu (2006). The impact of explicit deposit insurance on market discipline. DNB Working Papers 089, Netherlands Central Bank, Research Department.

Iyer, R. and M. Puri (2012). Understanding bank runs: The importance of depositor-bank relationships and networks. *American Economic Review 102*(4), pp. 1414–1445.

Jagtiani, J., G. Kaufman, and C. Lemieux (1999). Do markets discipline banks and bank holding companies? Evidence from debt pricing. *Emerging Issues* (Jun).

Jagtiani, J. and C. Lemieux (2001). Market discipline prior to bank failure. *Journal of Economics and Business 53*(2-3), pp. 313–324.

Janicki, H. P. and E. S. Prescott (2006). Changes in the size distribution of U.S. banks: 1960-2005. *Economic Quarterly*, pp. 291–316.

Javorcik, B. S. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review 94*(3), pp. 605–627.

Jobson, J. D. and B. Korkie (1980). Estimation for Markowitz efficient portfolios. *Journal of the American Statistical Association 75*(371), pp. 544–554.

Johnson, E. J., J. Hershey, J. Meszaros, and H. Kunreuther (1993). Framing, probability distortions, and insurance decisions. In C. Camerer and H. Kunreuther (Eds.), *Making Decisions About Liability And Insurance*, pp. 35–51. Springer Netherlands.

Johnston, R. B. and O. M. Nedelescu (2006). The impact of terrorism on financial markets. *Journal of Financial Crime 13*(1), pp. 7–25.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica 47*(2), pp. 263–291.

Karas, A., W. Pyle, and K. Schoors (2010). How do Russian depositors discipline their banks? Evidence of a backward bending deposit supply function. *Oxford Economic Papers 62*(1), pp. 36–61.

Karas, A., W. Pyle, and K. Schoors (2013). Deposit insurance, banking crises, and market discipline: Evidence from a natural experiment on deposit flows and rates. *Journal of Money, Credit and Banking 45*(1), pp. 179–200.

Karolyi, G. A. and R. Martell (2010). Terrorism and the stock market. *International Review of Applied Financial Issues and Economics 2*(2), pp. 285–314.

Kip Viscusi, W. (2009). Valuing risks of death from terrorism and natural disasters. *Journal of Risk and Uncertainty 38*(3), pp. 191–213.

Klibanoff, P., O. Lamont, and T. A. Wizman (1998). Investor reaction to salient news in closed-end country funds. *Journal of Finance 53*(2), pp. 673–699.

Laeven, L. and F. Valencia (2012). Systemic banking crises database: An update. IMF Working Papers 12/163, International Monetary Fund.

LaFree, G. and L. Dugan (2007). Introducing the global terrorism database. *Terrorism and Political Violence 19*(2), pp. 181–204.

Ledoit, O. and M. Wolf (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance 10*(5), pp. 603–621.

Ledoit, O. and M. Wolf (2004). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management 30*(4), pp. 110–119.

Levitt, S. D. (2004). Why are gambling markets organised so differently from financial markets? *Economic Journal 114*(495), pp. 223–246.

Lo, A. W. (2007). Efficient market hypothesis. In L. Blume and S. Durlauf (Eds.), *The New Palgrave: A Dictionary of Economics* (2 ed.). New York: Palgrave McMillan.

Loutskina, E. and P. E. Strahan (2011). Informed and uninformed investment in housing: The downside of diversification. *Review of Financial Studies 24*(5), pp. 1447–1480.

Lucas, R. E. (1978). Asset prices in an exchange economy. *Econometrica 46*(6), pp. 1429–1445.

Maddala, G. S. (1987). Limited dependent variable models using panel data. *Journal of Human Resources 22*(3), pp. 307–338.

Maechler, A. M. and K. M. McDill (2006). Dynamic depositor discipline in US banks. *Journal of Banking & Finance 30*(7), pp. 1871–1898.

Maheu, J. M. and T. H. McCurdy (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance 59*(2), pp. 755–793.

Maillet, B. B. and T. L. Michel (2005). The impact of the 9/11 events on the American and French stock markets. *Review of International Economics 13*(3), pp. 597–611.

Makropoulou, V. and R. N. Markellos (2011). Optimal price setting in fixed-odds betting markets under information uncertainty. *Scottish Journal of Political Economy 58*, pp. 519–536.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives 17*(1), pp. 59–82.

Malmendier, U. and S. Nagel (2011). Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics 126*(1), pp. 373–416.

Markose, S., S. Giansante, and A. R. Shaghaghi (2012). 'Too interconnected to fail' financial network of US CDS market: Topological fragility and systemic risk. *Journal of Economic Behavior & Organization 83*(3), pp. 627–646.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance 7*(1), pp. 77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons.

Marsh (2005). Marketwatch: Terrorism Insurance 2005.

Marsh (2010). The Marsh Report: Terrorism Risk Insurance 2010.

Martinez Peria, M. S. and S. L. Schmukler (2001). Do depositors punish banks for bad behavior? Market discipline, deposit insurance, and banking crises. *Journal of Finance 56*(3), pp. 1029–1051.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics: Economic theory and mathematical economics*, pp. pp. 105–142. Academic Press, New York.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), pp. 1695–1725.

Melnick, R. and R. Eldor (2010). Small investment and large returns: Terrorism, media and the economy. *European Economic Review 54*(8), pp. 963–973.

Mende, A. (2006). 09/11 on the USD/EUR foreign exchange market. *Applied Financial Economics 16*, pp. 213–222.

Merton, R. (1976). Option pricing when underlying stocks are discontinuous. *Journal of Financial Economics 3*, pp. 125–144.

Michel-Kerjan, E. O., P. A. Raschky, and H. Kunreuther (2011). Do firms manage catastrophe and non-catastrophe risks differently? The Wharton School Risk Management and Decision Processes Center - Working Paper # 2011-13.

Montalvo, J. G. (2011). Voting after the bombings: A natural experiment on the effect of terrorist attacks on democratic elections. *Review of Economics and Statistics 93*(4), pp. 1146–1154.

Mátyás, L. and L. Balázsi (2012). The estimation of multi-dimensional fixed effects panel data models. CEU Working Papers 2012/2, Department of Economics, Central European University.

Naik, V. (1993). Option valuation and hedging strategies with jumps in volatility of asset returns. *Journal of Finance 48*(5), pp. 1969–1984.

Office of the Special Inspector General for the Troubled Asset Relief Program (2013, October 29). Quarterly report to congres. Technical report, US Treasury.

Oliveira, R. d. F., R. F. Schiozer, and L. A. B. d. C. Barros (2014). Depositors' perception of "Too-Big-to-Fail". *Review of Finance*, forthcoming.

Onay, C. and E. Ozsoz (2013). The impact of Internet-banking on brick and mortar branches: The case of Turkey. *Journal of Financial Services Research 44*(2), pp. 187–204.

Papke, L. and J. Wooldridge (1996). Econometric methods for fractional response variables with an application to 401(K) plan participation rates. *Journal of Applied Econometrics 11*, pp. 619–632.

Papke, L. E. and J. M. Wooldridge (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics 145*, pp. 121–133.

Park, K. and G. Pennacchi (2009). Harming depositors and helping borrowers: The disparate impact of bank consolidation. *Review of Financial Studies 22*(1), pp. 1–40.

Park, S. (1995). Market discipline by depositors: Evidence from reduced-form equations. *Quarterly Review of Economics and Finance 35*, pp. 497–514.

Park, S. and S. Peristiani (1998). Market discipline by thrift depositors. *Journal of Money, Credit and Banking 30*(3), pp. 347–364.

Park, S. and S. Peristiani (2007). Are bank shareholders enemies of regulators or a potential source of market discipline? *Journal of Banking & Finance 31*(8), pp. 2493–2515.

Peren Arin, K., D. Ciferri, and N. Spagnolo (2008). The price of terror: The effects of terrorism on stock market returns and volatility. *Economics Letters 101*, pp. 164–167.

Piazza, J. A. (2008). A supply-side view of suicide terrorism: A cross-national study. *Journal of Politics 70*(1), pp. 28–39.

Poteshman, A. M. (2006). Unusual option market activity and the terrorist attacks of September 11, 2001. *Journal of Business 79*(4), pp. 1703–1726.

Prager, R. A. and T. H. Hannan (1998). Do substantial horizontal mergers generate significant price effects? evidence from the banking industry. *Journal of Industrial Economics 46*(4), pp. 433–452.

Quandt, R. (1986). Betting and equilibrium. *Quarterly Journal of Economics 101*, pp. 201–207.

Radecki, L. J. (1998). The expanding geographic reach of retail banking markets. *Federal Reserve Bank of New York Economic Policy Review 4*(2), pp. 15–34.

Rajan, R. G. (2005). Has financial development made the world riskier? *Proceedings Aug*, pp. 313–369.

Rios-Avila, F. (2013). Feasible estimation of linear models with N-fixed effects. Economics Working Paper Archive 782, Levy Economics Institute.

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. *Industrial management review 6*(2), pp. 41–49.

Schnytzer, A., M. Lamers, and V. Makropoulou (2010a). Measuring the extent of inside trading in horse betting markets. *Journal of Gambling Business and Economics 4*(2), pp. 21–41.

Schnytzer, A., M. Lamers, and V. Makropoulou (2010b). The impact of insider trading on forecasting in a bookmakers' horse betting market. *International Journal of Forecasting 26*(3), pp. 537–542.

Schnytzer, A., V. Makropoulou, and M. Lamers (2014). Pricing decisions and insider trading in horse betting markets. In L. Vaughan-Williams and D. Siegel (Eds.), *Handbook on the Economics of Gambling*. London: Oxford University Press.

Schnytzer, A. and Y. Shilony (1995). Inside information in a betting market. *Economic Journal 105*, pp. 963–971.

Schnytzer, A. and Y. Shilony (2002). On the timing of inside trades in a betting market. *European Journal of Finance 8*(2), pp. 176–186.

Schnytzer, A. and Y. Shilony (2003). Is the presence of insider trading necessary to give rise to a favourite-longshot bias? In L. V. Williams (Ed.), *The Economics of Gambling*, pp. 14–17. Routledge.

Schnytzer, A. and A. Snir (2008). Herding in imperfect markets with inside traders. *Journal of Gambling Business and Economics 2*(2), pp. 1–16.

Shin, H. S. (1991). Optimal betting odds against insider traders. *Economic Journal 101*(408), pp. 1179–85.

Shin, H. S. (1992). Prices of state contingent claims with insider traders, and the favourite-longshot bias. *Economic Journal 102*(411), pp. 426–35.

Shin, H. S. (1993). Measuring the incidence of insider trading in a market for state-contingent claims. *Economic Journal 103*(420), 1141–53.

Silver, N. (2012). *The Signal and the Noise: Why So Many Predictions Fail-but Some Don't*. Penguin Press.

Sironi, A. (2003). Testing for market discipline in the European banking industry: Evidence from subordinated debt issues. *Journal of Money, Credit and Banking 35*(3), pp. 443–72.

Snowberg, E. and J. Wolfers (2010). Explaining the favorite-long shot bias: Is it risk-love or misperceptions? *Journal of Political Economy 118*(4), pp. 723–746.

Stiroh, K. J. (2004). Diversification in banking: Is noninterest income the answer? *Journal of Money, Credit and Banking 36*(5), pp. 853–882.

Stiroh, K. J. and A. Rumble (2006). The dark side of diversification: The case of US financial holding companies. *Journal of Banking & Finance 30*, pp. 2131–2161.

Straetmans, S. T. M., W. F. C. Verschoor, and C. C. P. Wolff (2008). Extreme US stock market fluctuations in the wake of 9/11. *Journal of Applied Econometrics 23*(1), pp. 17–42.

Study of Terrorism and Responses to Terrorism (START) (2011). Global Terrorism Database: GTD variables & inclusion criteria.

Sunstein, C. (2003). Terrorism and probability neglect. *Journal of Risk and Uncertainty 26*(2-3), 121–136.

Taleb, N. N. (2007). *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets* (2 ed.). London: Random House.

Thompson Jr, W. C. (2002). One year later: The fiscal impact of 9/11 on New York City. *City of New York Office of the Comptroller, New York (4 September)*.

Torres, S., P. Portugal, J. T. Addison, and P. Guimaraes (2013). The sources of wage variation: A three-way high-dimensional fixed effects regression model. IZA Discussion Papers 7276, Institute for the Study of Labor (IZA).

Tversky, A. and D. Kahneman (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology 5*(2), pp. 207–232.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), pp. 1124–1131.

United Nations Conference on Trade and Development (UNCTAD) (2011). World Investment Report 2011.

van Lelyveld, I. and F. Liedorp (2006). Interbank contagion in the Dutch banking sector: A sensitivity analysis. *International Journal of Central Banking 2*(2), pp. 99–133.

Vaughan Williams, L. (1999). Information efficiency in betting markets: a survey. *Bulletin of Economic Research 51*(1), pp. 1–39.

Vaughan Williams, L. and D. Paton (1997). Why is there a favourite-longshot bias in british racetrack betting markets. *Economic Journal 107*(440), pp. 150–158.

Wachter, J. A. (2013). Can time-varying risk of rare disasters explain aggregate stock market volatility? *Journal of Finance 68*(3), pp. 987–1035.

Wagner, W. (2008). The homogenization of the financial system and financial crises. *Journal of Financial Intermediation 17*(3), pp. 330–356.

Wheelock, D. C. and P. W. Wilson (2012). Do large banks have lower costs? New estimates of returns to scale for U.S. banks. *Journal of Money, Credit and Banking 44*(1), pp. 171–199.

Zhou, C. (2010). Are banks too big to fail? Measuring systemic importance of financial institutions. *International Journal of Central Banking 6*(4), pp. 205 – 250.

Nederlandse samenvatting

Dit proefschrift bestudeert gedragsveranderingen van economische actoren in financiële beslissingen als gevolg van extreme gebeurtenissen. Hierbij staan drie vragen centraal. Ten eerste, hoe houden economische actoren ex-ante rekening met extreme gebeurtenissen in financiële beslissingen? Ten tweede, hoe veranderen extreme gebeurtenissen het gezamelijke gedrag van economische actoren ex-post? Ten derde, hoe kunnen we de gevolgen van extreme gebeurtenissen minimaliseren? In tegenstelling tot voorgaande literatuur, welke deze gedragsveranderingen analyseren door middel van experimentele technieken, worden de veranderingen in dit proefschrift geanalyseerd aan de hand van geaggregeerde (prijs) data.

Prijsbesluiten, voorspellingen en extreme gebeurtenissen

Het tweede en derde hoofdstuk beantwoorden de vraag hoe economische actoren ex-ante rekening houden met extreme gebeurtenissen in financiële beslissingen. In deze hoofdstukken wordt het gedrag van een market maker geanalyseerd die wordt geconfronteerd met twee soorten tegenpartijen: handelaren met gewone kennis (noise traders) en handelaren met voorkennis (insiders). Deze laatste groep heeft superieure kennis vergeleken met de market maker, waardoor deze zijn prijzen zal aanpassen om zo min mogelijk geld te verliezen aan de insiders. De markt die wordt geanalyseerd in de hoofdstukken is een gok-markt waar de market maker weddenschappen accepteert tegen vaste kansen (ook wel odds genoemd). In deze markt wordt geld ingezet op een weddenschap met een duidelijk eindpunt. Deze eigenschappen maken de gok-markt dan ook extreem geschikt om hypothesen te testen die oorspronkelijk zijn ontwikkeld voor financiële markten (Vaughan Williams, 1999). Omdat de market maker (in gokmarkten ook wel bookmaker genoemd) grote posities inneemt op de uitkomst van de weddenschap (Levitt, 2004), wil deze geen geld verliezen aan gokkers met superieure kennis van zaken. Hoofdstuk 2 ontwikkelt een model waarin een bookmaker prijzen moet vaststellen wanneer deze wordt geconfronteerd met de twee soorten gokkers. De assumptie hierbij is dat insiders een betere kennis van zaken hebben dan de bookmaker met betrekking tot de werkelijke kansen, wat in de lijn ligt met de Bayesiaanse gokker in Silver (2012). Echter, deze definitie is anders dan de voorgaande literatuur, welke uitgaat van insiders die vantevoren weten wie er gaat winnen (Shin, 1991, 1992, 1993). Het hoofdstuk laat zien dat de prijzen die worden vastgesteld door de bookmaker gelijk zijn aan een (reële) call-optie die wordt aangeboden aan de insider: hoe groter het verschil tussen de prijzen van de bookmaker en de verwachting van de insider, hoe hoger de kans dat de insider deze optie zal uitvoeren. Deze dynamiek is het belangrijkst bij weddenschappen op longshots (een kleine kans op winnen met een mogelijk hoge uitbetaling), aangezien de bookmaker hier meer geld op dreigt te verliezen dan op favorieten (een hoge kans op winnen met een lage uitbetaling). Het model voorspelt dat door de aanwezigheid van insiders, bookmakers hun odds zullen aanpassen om weddenschappen op longshots minder aantrekkelijk te maken voor insiders en hun mogelijke verlies te minimaliseren. Hiermee is het model consistent met, en geeft het een mogelijke oorzaak voor de favorite-longshot bias in gok-markten (Vaughan Williams en Patton, 1997; Snowberg en Wolfers, 2010). Deze bias wordt geconstateerd wanneer de kansen die worden geïmpliceerd door de prijzen van de bookmaker onjuist zijn, omdat favorieten (longshots) een consistent lagere (hogere) impliciete winstkans hebben dan achteraf blijkt. Door het maken van bepaalde assumpties over hoe nieuws dat relevant is voor de weddenschappen zich ontwikkelt, is het mogelijk om de call-opties die aan de insider worden aangeboden te simuleren. Hierna wordt het model gekalibreerd op bookmaker odds die zijn verzameld in Australische paarden-races in 1998. Uit de prijzen kan worden opgemaakt dat de bookmaker verwacht dat 97% van het ingezette geld van insiders komt. Wanneer dit wordt gecombineerd met objectieve (ex-post) observaties over insider trading, komt ruwweg 60% van de weddenschappen van insiders en 40% van noise traders. De bookmaker overschat in grote mate de kans dat hij verlies zal maken door de aanwezigheid van insiders, wat consistent is met het niet juist kunnen inschatten van de kans dat extreme gebeurtenissen plaatsvinden (Tversky en Kahneman, 1973). Hoofdstuk 3 bekijkt verder of deze call-opties extra informatie bevatten vergeleken met de prijzen – in welke alle publiek verkrijgbare data zijn verwerkt. Hoewel er, statistisch gezien, extra verklarende kracht voortkomt uit de call-opties, is de economische impact klein aangezien er op basis van deze call-opties geen winstgevende trading rule kan worden gevonden.

Gedrag van investeerders en extreme gebeurtenissen

Hoofdstuk 4 analyseert hoe extreme gebeurtenissen het gedrag van economische actoren kan beïnvloeden *na* het plaatsvinden. Eerder onderzoek heeft aangetoond dat extreme gebeurtenissen dit gedrag op zowel korte- als lange-termijn kan veranderen. Eén extreme gebeurtenis die vaak bestudeerd wordt, is de terroristische aanslag van 11 september op het World Trade Center in New York, en met name de invloed op financiële markten.[1] Over het algemeen wordt gevonden dat aandeelprijzen reageren op extreme gebeurtenissen zoals terrorisme (Chen and Siems, 2004; Drakos, 2004; Eldor and Melnick, 2004; Karolyi and Martell, 2010; Chesney et al., 2011), maar deze reacties verschillen per aanval, per sector van de economie en per land. Hoewel het is vastgesteld dat terroristische aanvallen kunnen leiden tot reacties van aandeelprijzen in een ander land, is het nog onduidelijk via welk kanaal deze schokken zich bewegen. Abadie en Gardeazabal (2003) en Drakos (2010a) hebben recent geopperd dat de schokken zich voortbewegen door middel van de economische verbanden – zoals handel – tussen landen. Deze studies gebruiken echter alleen aanvallen vóór of ná 11 september 2001, en voeren geen vergelijkende studie uit wat de invloed van 9/11 op dit verband was. Hoofdstuk 4 voert deze vergelijkende studie wel uit, en bevestigt dat economische verbanden inderdaad kunnen verklaren hoe verschillende landen reageren op aanvallen in het buitenland. De analyse op Amerikaanse aandeelprijzen laat zien dat deze prijzen reageren op terroristische aanslagen buiten Amerika. Daarnaast zijn deze reacties inderdaad evenredig aan de hoeveelheid investeringen van Amerikaanse bedrijven in deze regio: aanvallen in regio's met meer Amerikaanse investeringen leiden tot een grote reactie in de aandeelprijzen. In de vergelijkende studie vóór en ná 11 september 2001, wordt gevonden dat het kanaal alleen statistisch significant is *na* 9/11. Na vervolgonderzoek is de meest aannemelijke verklaring dat de Amerikaanse beleggers vóór 9/11 leiden aan kortzichtigheid met betrekking tot terroristische aanvallen

[1] Voor een uitgebreid overzicht over de invloed van 9/11 op financiële markten en andere terreinen verwijs ik naar Frey and Luechinger (2005).

omdat ze hier weinig aan waren blootgesteld. Ná 9/11 reageren de beleggers wel op aanvallen in het buitenland en doen dit meer voor aanvallen in gebieden waar Amerikaanse bedrijven meer investeringen hebben. Dit duidt erop dat de aanslagen van 11 september 2001 een wake-up call was voor beleggers, met als gevolg een groter besef voor de gevolgen van terrorisme in het buitenland voor Amerikaanse bedrijven.

Gedrag van deposito-houders, bank grootte en extreme gebeurtenissen

De laatste twee hoofdstukken van dit proefschrift zijn geïnspireerd door de extreme gebeurtenissen in de bankensector en financiële sector in de Verenigde Staten in 2007 en 2008. De financiële crisis die het gevolg was van deze extreme gebeurtenissen leidde tot het faillissement van grote financiële instellingen, bijvoorbeeld de zakenbank Lehman Brothers, en bedreigde de veiligheid van het financiële bestel. Hoewel overheden over de hele wereld moesten bijspringen om de grootbanken overeind te houden, gingen kleine banken wél failliet. In de Verenigde Staten bijvoorbeeld zijn er ongeveer 400 commerciële banken gesloten door de Federal Deposit Insurance Corporation (FDIC) sinds het begin van de crisis, en er zijn nog eens 500 noodlijdende banken overgenomen door gezondere banken. In Hoofdstuk 5 worden de effecten van deze faillissementen op spaarders bekeken. Spaarders hebben een rol in het toezicht op banken en kunnen deze 'straffen' door het opnemen van hun tegoeden wanneer ze de banken ervan verdenken te risicovol om te gaan met hun spaargeld. De reddingsoperaties en bankfaillissementen kunnen spaarders gemengde signalen geven over de noodzaak om toe te zien op de activiteiten van hun banken. Aan de ene kant kunnen reddingsoperaties een verminderde prikkel geven aan spaarders, omdat deze er dan van uit gaan dat andere banken ook zullen worden geredt ongeacht hoe risicovol ze zijn. Aan de andere kant heeft voorgaande literatuur vastgesteld dat faillissementen kunnen leiden tot extra toezicht van spaarders, wat veroorzaakt wordt door een hogere risico-aversie (Martinez Peria en Schmukler, 2001; Karas et al., 2010, 2013; Iyer en Puri, 2012). Omdat er tijdens de crisis zowel reddingsoperaties als bankfaillissementen hebben plaatsgevonden, is het onduidelijk welke van deze effecten zal domineren in het gedrag van spaarders. Hoofdstuk 5 stelt vast dat in markten met faillissementen, spaarders meer toezicht houden op banken ondanks de reddingsoperaties. Daarnaast is er meer toezicht in regio's met meerdere faillissementen.

Hoofdstuk 6 probeert lessen te trekken uit de recente financiële crisis met als doel om een nieuwe crisis te voorkomen. De vele reddingsoperaties hebben de overheidsfinanciën flink belast en veel nieuwe voorstellen zijn gedaan om dit in de toekomst te vermijden. De nadruk in deze voorstellen ligt vaak op de onderlinge verbondenheid van de grootbanken, de mate van inkomensdiversifiëring, of de omvang van banken (De Jonghe, 2010; Zhou, 2010; Adrian en Brunnermeier, 2011; Markose et al., 2012; Bertay et al., 2013). Hoofdstuk 6 bekijkt deze crisis uit het oogpunt van de toezichthouder als beheerder van een portefeuille van banken, en bestudeert de optimale portefeuille die de toezichthouder zou willen hebben om zoveel mogelijk het systeem risico te verminderen. De bevindingen zijn dat de concentratie in de huidige portefeuille consequent te hoog is en dat een minder geconcentreerde banken-portefeuille een veel lager risico met gelijkaardige opbrengsten zou hebben.

Suggesties voor vervolgonderzoek

De hoofdstukken zijn verbonden door gedrag van economische actoren dat afwijkt van rationeel gedrag, veroorzaakt door het niet juist kunnen inschatten van de kans dat extreme gebeurtenissen plaatsvinden (Tversky en Kahneman, 1973; Kahneman en Tversky, 1979). Dit proefschrift laat zien dat dit gedrag zich ook voordoet wanneer er wordt gekeken naar geaggregeerde (prijs) data. Zodoende verschaft het bewijs dat de standaard modellen in de economie, die uitgaan van rationeel gedrag van de populatie van actoren, onjuist zijn en dat de actoren samen ook irrationele beslissingen nemen als gevolg van extreme gebeurtenissen. Toekomstig onderzoek zal hier rekening mee moeten houden en zich moeten focussen op de groepsdynamiek van financiële beslissingen. Daarnaast is er ook meer onderzoek nodig naar ander bewijs van irrationeel gedrag in geaggregeerde (prijs) data naar aanleiding van andere gebeurtenissen dan extreme gebeurtenissen.

Curriculum Vitae

Martien Lamers (1985) was born in Nieuwegein, the Netherlands. After a B.Sc. in Economics & Law, he obtained a M.Sc. in International Economics & Business from Utrecht University in 2008, graduating *summa cum laude*. His master thesis, entitled “Insider Trading in Betting Markets”, was nominated by the Utrecht School of Economics for the yearly thesis award of Utrecht University and forms the basis for the first two chapters of this dissertation. Martien obtained an M.Sc. in Banking & Finance from Ghent University in 2009, graduating *magna cum laude* and was awarded the Europa Bank award for best Banking & Finance student in his year. He wrote his master thesis on the default risk of small and medium enterprises during an internship at Deloitte Financial & Actuarial Risk Management.

In 2009, Martien joined the Department of Financial Economics at Ghent University as a Ph.D. student, obtaining a 4 year grant for his doctoral dissertation proposal from the Special Research Fund (Bijzonder Onderzoeksfonds) in 2010. As a teacher, he supervised several master theses and taught courses on Investment Analysis, Financial Risk Management and Topics in Empirical Research in Finance.

Martien’s papers have been published in the International Journal of Forecasting, the Oxford Handbook on the Economics of Gambling, and Bank- en Financiewezen. He has presented his research at international conferences organized by, amongst others, the Canadian Economics Association, the Royal Economic Society, the European Economic Association, the Financial Management Association and the Financial Engineering and Banking Society. Moreover, Martien has been on research visits at Maastricht University and Utrecht University, where he was also an invited guest lecturer.

In August 2014, Martien joined the Faculty of Economics, Econometrics and Finance at the University of Groningen as a tenure track assistant-professor. His research interests are the impact of extreme events in banking, financial stability and gambling.