UNIVERSITÀ COMMERCIALE "LUIGI BOCCONI"

PHD SCHOOL

PhD program in: Economics and Finance

Cycle: 35

Disciplinary Field (code): SECS-P/09

Efficiency in Firm Financing and Financial Markets

Advisor: Stefano Rossi

Co-Advisor: Mariano Massimiliano Croce

PhD Thesis by

Alireza Aghaee Shahrbabaki

ID number: 3112008

Year: 2025

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my advisors, Stefano Rossi and Max Croce. Their unwavering guidance, intellectual generosity, and steadfast support have been a cornerstone of my academic journey. I am profoundly thankful for the time, energy, and wisdom they devoted to nurturing my research. They offered me the perfect balance of direction and independence—providing the knowledge and structure to navigate my path, while allowing me the freedom to grow, explore, and shape my own scholarly voice. Their mentorship has left a lasting imprint on both my professional and personal development, and I will remain forever indebted to them.

I am also sincerely grateful to Lorenzo Bretscher for his kind support, particularly during the pivotal stages of my job market season. His encouragement and advice have been invaluable.

My heartfelt thanks extend to Dimitri Vayanos for graciously hosting me during my semester at the London School of Economics and Political Science. The stimulating conversations with Dimitri were a source of inspiration and intellectual renewal, pushing me to refine my thinking and aspire to higher standards. I am also thankful to Ricardo Colacito for hosting me as a visiting PhD student at the University of North Carolina, Kenan-Flagler Business School and for his support and advice.

To my family, I owe a debt of gratitude beyond words. Your unwavering love, encouragement, and belief in me have been the bedrock of my strength. A special thank you to my parents, whose faith in my potential and enduring support have been a constant light through every challenge and milestone.

A heartfelt thank you to my dearest friend, Amir, who has stood by me through the highs and lows of this journey. His steady presence, willingness to listen, and unshakable friendship—regardless of time zones or distance—have been a source of comfort, strength, and joy. Our bond has only grown deeper through the years, and I cherish it deeply.

Finally, I extend my heartfelt appreciation to the finance faculty, colleagues, and friends at Bocconi University. I am especially grateful to Felix and Giuliano, whose friendship and support have meant a great deal to me throughout this journey. The vibrant intellectual community, the spirit of companionship, and the many enriching discussions have made my time here profoundly rewarding. I am deeply thankful for the friendships and collaborations that have shaped not only my academic experience but also made this journey truly meaningful.

Abstract

Efficiency is a central tenet of classical finance theory, grounded in the assumption that agents act to eliminate profit opportunities. Real-world markets often deviate from these idealized predictions due to frictions. My dissertation examines such frictions from both corporate and investor perspectives, focusing on their origins, deviations from theoretical benchmarks, and broader impact on financial markets.

My first paper explores how information-free demand shocks affect stock prices. Under the efficient market hypothesis, stock prices reflect expected future cash flows discounted for systematic risk, implying nearly flat demand curves—meaning that non-informational demand shocks should not move prices. Yet, demonstrating this empirically requires isolating truly information-free shocks.

One prominent setting is the "index effect," where stocks added to or removed from major indices exhibit abnormal returns. If these changes contain no new fundamental information, they challenge the idea of flat demand curves and thus the efficient market hypothesis. However, as the magnitude of the index effect has declined over time despite growing passive investing flows, scholars have debated whether this reflects declining inefficiencies or information in index reconstitutions.

To address this, I propose a novel identification approach that shifts focus from the firms added or removed during reconstitution to those index members not directly involved in these events. Since the sizes of added and deleted firms generally differ, the portfolio weights of incumbent firms must adjust to keep the total at 100%, requiring that passive

index trackers alter their demand for incumbent stocks without any new information about them. Consequently, any abnormal returns on these stocks around reconstitution can be linked solely to passive demand shifts. By studying these information-free demand shifts on incumbents' prices, I can estimate what the abnormal returns would be for added and deleted stocks in the absence of an informational component. Comparing the realized abnormal returns on additions and deletions with these counterfactual scenarios enables me to isolate the role of information.

By estimating counterfactual abnormal returns for added and deleted stocks based on incumbent behavior, I isolate the informational component of the index effect. I find that post-2000, the index effect is largely driven by passive demand, with information playing a minor role. Specifically: (i) a simple demand-based counterfactual—accounting for time-varying elasticities—can replicate the average effect size and its decline over time; and (ii) the declining effect stems from increasingly flatter demand curves, not improved information efficiency. In contrast, pre-2000 effects were too large to be explained by passive demand alone, suggesting a stronger role for information in earlier periods.

My second paper, coauthored with Stefano Rossi and Lorenzo Bretscher, investigates how legal environments shape firms' financing strategies and creditors' responses. Using newly combined datasets, we reconstruct the debt structure of 10,136 firms across 51 countries. We find that debt ownership is most concentrated in civil law countries and most dispersed in common law countries.

In strong investor protection environments, firms borrow through both dispersed unsecured debt and concentrated secured bank lending. Where investor protection is weaker, firms respond by borrowing shorter-term and in USD-denominated debt. These patterns are most pronounced among small and medium-sized firms, while large firms tend to attract dispersed debt from international investors—effectively circumventing local institutional constraints. Our work illustrates how legal environments influence firm behavior and how creditors strategically structure ownership to manage the trade-off between strategic default and inefficient liquidation.

Contents

Chapter 1

The Flattening Demand Curves

Abstract

Abnormal returns for stocks added to or removed from the S&P 500 index (known as index effect) have been declining, despite a sharp rise in demand shifts of passive funds. I explore (i) whether these abnormal returns during index reconstitutions stem from passive demand shifts or information content, and (ii) why they have decreased over time. My study isolates pure demand-based price impact using a novel identification strategy that analyzes incumbent stocks–index constituents whose portfolio weights adjust due to the differing sizes of added and deleted firms. I find that (i) the index effect was primarily driven by information prior to 2000 while passive demand became predominant thereafter, and (ii) the declining index effect is a direct result of the flattening of stocks' demand curves in the context of index reconstitutions, independent of the informativeness of the index committee's decisions. This flattening is associated with decreased arbitrage risk and enhanced cross-stock substitution.

Keywords: Index Effect, Elasticity of Demand, Market Efficiency, Index Reconstitution, Index Funds, Arbitrage Risk.

1.1 Introduction

The literature has documented the existence of an *index effect*, that is, index reconstitutions are associated with abnormal returns experienced by stocks added to or removed from a major equity index (Shleifer, 1986; Harris and Gurel, 1986; Pavlova and Sikorskaya, 2022). If additions and deletions do not convey any fundamental information, the index effect challenges the efficient market hypothesis, according to which stocks' demand curves should be flat (Petajisto, 2009). Consequently, the literature has inferred that either index changes convey information (Denis et al., 2003; Chen et al., 2004; Cai, 2007) or that market efficiency is compromised, leading to downward-sloping demand curves for stocks (Shleifer, 1986; Harris and Gurel, 1986). This latter option has become increasingly controversial in the last two decades; despite the significant growth of passive investing, there has been a secular decline in the magnitude of the index effect in recent decades (Greenwood and Sammon, 2024). In an economy with a downward-sloping demand curve, one would expect larger price reactions if demand shocks were increasing in magnitude. Using a novel identification strategy, this study revisits the index effect literature for S&P 500 in search of the origin of observed abnormal returns, their determinants, and the reasons for their decreasing trend using a sample of index portfolio from 1990 to 2021. My findings show that demand is the primary driver of index effect for S&P 500 post-2000, and the declining index effect is a direct result of the flattening of stocks' demand curves, which is related to decreased risk of arbitrage.

In the existing literature, the demand-based explanation of the index effect attributes abnormal returns to the inelastic demand of passive index followers who must buy added firms and sell deleted firms regardless of their prices. Conversely, the information-based explanation posits that index events convey new information to the market, which in turn moves the prices. The S&P 500 index is an ideal candidate to disentangle demand- versus information-based price adjustments because its composition changes are not rule-based and even do not follow a fixed schedule. In the case of S&P family of indices the index committee, composed of financial experts with first-hand information, makes all decisions, thereby rendering their determinations indicative of relevant underlying information in the eyes of the market[1]. Thus, investors might conclude that index decisions convey information. This cannot be said about other indices that are rule-based and, in a way, mechanical. This is also why most of the evidence in the literature related to any form of information revolves around the S&P family.

To assess the relative explanatory power of demand- and information-based theories in explaining the index effect, I propose a novel identification that shifts the focus from the added and deleted firms to other index members that were not actively involved in the reconstitution events. These firms are referred to as index incumbents, as they were index constituents both before and after the reconstitution event. Since the size of added firms typically differs from that of deleted firms, the portfolio weight of the incumbent firms must change as well. This is purely mechanical so that the sum of all weights remains equal to 100%. Most importantly, passive index trackers must adjust their demand of incumbent stocks even though no additional information has been conveyed about them. Therefore, any abnormal return on these stocks around the reconstitution can be associated with passive demand shifts. Studying the effect of such information-free demand shifts on incumbents' prices enables me to predict what would be the abnormal returns for the added and deleted stocks if there was no informational component in the index effect. By comparing realized abnormal returns observed on additions and deletions with the counterfactual demand-driven scenarios of my model I can disentangle the role of information. I describe my findings below.

[1]Different stock market indices use varying methods and criteria to select and weigh their constituent stocks. Some indices, like the FTSE and NASDAQ, are fully transparent and rule-based, making it relatively manageable to predict which stocks will be included or excluded (Danbolt et al., 2018). However, other indices like the Russell and CRSP are also rule-based and transparent, but their changes are not entirely predictable (Chang et al., 2014; Heath et al., 2020). Finally, there are indices like the S&P family, whose inclusion or exclusion of stocks is decided by a committee and is entirely subjective, lacking even a fixed calendar.

First, In the post-2000 period, my findings show that passive demand is the main driver of the index effect and that information has a negligible role. I show: (i) a simple demand-based counterfactual exhibiting time-varying elasticities estimated on index incumbents can reproduce the entire average effect size on additions and deletions and their pattern over time, and (ii) the decreasing pattern of index effect is due to the flattening of stocks' demand curves in the context of index reconstitutions, which extends beyond and is unrelated to index decisions informativeness. In the pre-2000 era, however, I show that the index effect was too large to be reasonably attributed only to the demand shifts of index funds given their small size relative to firms' market capitalization, and there was indeed a primary information-based component in the index effect.

Second, I have a methodological contribution related to my novel identification strategy to disentangle the price effects due to a pure demand shift from those caused by informative news. Previous studies focused on the price impact of additions and deletions which, by their nature, embody both demand- and information-based effects. By shifting the focus toward incumbents, I can focus on information-free price changes. More specifically, I derive a new measure that captures the aggregate inelastic demand shift of all passive index trackers. I denote it as Mechanical Rebalancing Flow (MRF). We can think of the MRF as the amount of money pumped into or withdrawn from stocks via passive funds purely due to the adjustment in index composition relative to the stock's market capitalization.

Third, I highlight that MRF is a valid instrument for demand shifts, because it is positively associated with the price reaction, volume of trades, and volatility of prices around the index reconstitutions. The correlation between returns and demand shifts, however, start decreasing after 2000, implying that while the demand curves for stocks generally slope downward, they have significantly flattened in recent decades. This critical finding of the paper is key to addressing the puzzling phenomenon of the shrinking index effect magnitude despite the surge in passive investing, and hence, the size of underlying demand shifts.

I show that in my sample, holding all other factors constant, a demand shock equivalent to one percent of the total shares outstanding generates a price movement of roughly 40 basis points, implying a micro elasticity of index funds' demand shifts in the order of -2.5, which is in line with recent empirical studies (among others, Pavlova and Sikorskaya (2022)), but implies far less elastic demand curves than those described in theoretical studies relying on the efficient market hypothesis (for instance, Petajisto (2009))[2]. However, the important finding of the current study is not to assert that we are on average far from the benchmark efficiency-which wouldn't be the first time in the literature-but instead to provide insights into how far we were, how it evolved over time, and where we are now. In theory, we think of arbitrageurs as the bright rational agents in the market responsible for restoring efficiency. In this context, however, I show that the evolution pattern is not supporting a monotonic story of improved efficiency in financial markets. Instead, I demonstrate that approaching the efficient benchmark only started around the turn of the century, possibly when the inefficiency became large enough to attract sufficient arbitrage capital towards it.

Fourth, related to the discussion above, I focus on arbitrage activity as the channel moderating the effect of demand shifts on prices. Following Wurgler and Zhuravskaya (2002), I measure risk of arbitrage as the variance of residuals of regressing one stock's excess return on that of a group of suitable substitutes. When the residuals are sizable (small), the stock taken in consideration can (cannot) be easily substituted by other stocks, that is, it has (does not have) close substitutes. I show that (i) prices of stocks with close substitutes are less sensitive to demand shifts, and (ii) close substitutes have increased over my sample. Equivalently, the flattening of the aggregate demand curves for stocks is associated with reduced arbitrage risk of stocks. In addition, I document that

[2]See Wurgler and Zhuravskaya (2002) for a relatively older review of elasticity estimates and Gabaix and Koijen (2022) for a review of the most recent estimates. See also Li and Lin (2022) for a discussion linking the theory predictions to empirical estimates.

both arbitrage risk and price impact multiplier have an important cyclical component related to the overall capability of the market to provide liquidity. Looking at deviation from trend, we can state that both figures spike during episodes of financial distress such as the 2008 financial crises and the COVID-19 pandemic. Related literature and further contributions of my study are discussed in what follows.

Related literature. This paper naturally relates to the vast literature on index effect. The closest empirical works to this research are Pavlova and Sikorskaya (2022), and Greenwood and Sammon (2024). Pavlova and Sikorskaya (2022) investigates the effects of stock membership in multiple indices on funds' demand, prices, and expected returns using an extensive group of US equity indices, with an additional emphasis on the role of active funds in these effects. My study differs in several dimensions. First, the exogenous source of variation in their study is stocks switching between Russell 1000 and Russell 2000, which is a rule-based index. Such rule-based switches leave no room for information-based explanations, and all their effects are correctly attributed to demand shifts in their study. As a result, they do not mention the role of information-based explanations, which are relevant for subjective indices such as the S&P 500 and is a contribution of the current study. Second, I establish that also in the case of S&P 500 as a subjective index, in the post-2000 era the average magnitude of index effect can be almost entirely explained by a demand-based model featuring time-varying price impact sensitivity; no other information-based explanation needed. Lastly, I link the flattening of demand curves to the risk of arbitrage activity through cross-stock substitution.

In a contemporaneous contribution, Greenwood and Sammon (2024) explores four reasons behind the decreasing magnitude of the index effect for S&P 500 and concludes that the decline is akin to other anomalies that tend to diminish once they become well-known. My study complements the picture in several dimensions. First, in their study, they do not account for the information conveyed through S&P 500 additions and deletions, which has been documented directly for the pre-2000 period (Cai, 2007; Denis et al., 2003) and is reconfirmed indirectly in this study as the residual of the total effect minus what can be reasonably attributed to mechanical demand shifts. The vanishing informativeness of index changes, is part of the reason for the disappearance of index effect. Second, the most promising explanation in their study is improved liquidity provision in the market, which is, in essence, not so different from the improved efficiency in my study. Liquidity and efficiency are two sides of the same coin. Yet, their evidence is mostly suggestive, such as increased dedicated trading desks for index trading in large institutions and little change in the overall institutional holdings of additions and deletions pre- and post-events. This study complements the picture they started drawing by introducing an identification that proves the market has become more efficient in the context of index reconstitutions, and this is made possible by studying incumbents rather than additions and deletions, for which we had not yet rejected the informational changes.

The literature on the index effect offers various theories to explain the size, permanency, and symmetry of the effect for additions and deletions. See Afego (2017) for an extensive review of this literature. Demand-based theories attribute the index effect to the large and inelastic demand from passive index trackers (see, among others, Shleifer 1986; Harris and Gurel 1986; Mase 2007; Beneish and Whaley 1996; Lynch and Mendenhall 1997; Kaul et al. 2000; Wurgler and Zhuravskaya 2002; Chang et al. 2014; Fernandes and Mergulhão 2016; Danbolt et al. 2018; Pavlova and Sikorskaya 2022).

Information-based theories propose that index reconstitutions somehow change the market expectation of the involved stock and thereby communicate new information about firms' prospects. Denis et al. (2003) examines the informational content of S&P 500 additions by investigating analysts' earnings forecast revisions around the index event and comparing post-inclusion realized earnings to pre-inclusion forecasts. They find that even stock analysts take information from index additions and exhaustively revise upward their initial earnings forecasts, and such revisions indeed improve their forecasts. Cai (2007) examines the information content of S&P 500 index changes by looking at the price and volume reaction of industry- and size-matched firms. He finds that the information content of index additions is so large that it even spills over to the industry-matched firms. Both studies use pre-2000 data and suggest that index reconstitutions convey information. Contrary to theirs, my analysis shows that S&P 500 additions and deletions no longer seem to convey information in the recent era. However, for the sample period they used, I can also verify the existence of such effects indirectly by studying the difference between actual and counterfactual information-free index effects. Furthermore, I directly replicate part of the study in Cai (2007) and show that the mechanical demand shift effects documented here are independent of their information effect.

The investor attention hypothesis suggests that heightened visibility spurs improved management performance (Denis et al., 2003). However, in a recent study, Bennett et al. (2023) demonstrates that the greater public scrutiny could also hurt a firm's performance. They show that elevated attention after index inclusion is often associated with declining post-inclusion performance. My study confirms their results and it expands their analysis to index deletions. I find that performance also improves for discretionary deletions, that are, deletions chosen by the index committee, as opposed to those that will be delisted soon after removal.

The information-based framework also features an awareness hypothesis, which posits that the inclusion of a firm in an index enhances investors' awareness of its existence. The investor awareness hypothesis justifies asymmetric effects observed on additions and deletions in their size and permanency. As Chen et al. (2004) puts it, "while more investors become aware of stocks added to the index, the number of investors aware of deleted stocks may not actually fall because it may be difficult for investors to become 'unaware' of those stocks." The finding in this study is motivated by the observation that the abnormal returns for deletions revert very soon in their sample. Similar to Chen et al. (2004), my findings reveal that the average cumulative abnormal return for additions stabilizes at a positive level after the event. Furthermore, I show that the reversion doesn't happen so fast for the discretionary deletions when they are distinguished from forced deletions, in which I define forced deletions as those which are entirely delisted from their exchange within two week from the index removal.

Last group of information-based explanations are related to liquidity. The liquidity hypothesis posits that increased analyst attention towards newly added firms enhances information production, diminishes information asymmetry, and ultimately boosts liquidity. This improved liquidity leads to lower required returns and higher prices for the newly added stocks (Amihud and Mendelson, 1986; Chen et al., 2004). Regarding the supply side of liquidity dynamics during index reconstitutions, Sikorskaya (2023) illustrates that the pricing impact of demand shifts can be moderated for stocks with limited availability of shares for short-selling. In light of the findings by Sikorskaya (2023), it is now challenging to classify liquidity effects solely among information-based explanations, as previously suggested in Afego (2017), since they also moderate the demand effect. However, this distinction does not alter the core conclusion of this paper, which asserts that a combination of factors unrelated to demand shifts influenced the S&P 500 index effect prior to 2000, a phenomenon that is no longer observed afterward. Whether the lending supply is classified as information-based or demand-based does not affect this conclusion.

This paper is also related to the literature on limits of arbitrage (see, among others, Shleifer and Vishny 1997; Wurgler and Zhuravskaya 2002; Greenwood 2005; Gromb and Vayanos 2010). Arbitrageurs in this context are rational agents who exploit mispricing opportunities until they diminish, accounting for the costs, constraints, and risks they encounter. Of particular relevance is Wurgler and Zhuravskaya (2002), who demonstrates that substitutability explains the cross-sectional price impact of demand shifts. In a contemporaneous study focusing on salient characteristics of corporate bonds such as rating and maturity, Chaudhary et al. (2022) also emphasize the significance of substitution in accurately determining cross- and self-elasticities of assets' demand. To the best of the author's knowledge, this study is the first to analyze the time trend of arbitrage risk for stocks, the emerging aggregate substitution patterns, and their connection to the evolution of demand elasticity over time.

There is also a vibrant literature on demand-system asset pricing that integrates downward-sloping demand curves for stocks into asset pricing and macro-finance models (see Koijen and Yogo 2019; Koijen et al. 2023; Gabaix and Koijen 2022; Gabaix et al. 2023). Koijen and Yogo (2019) link investors' preferences for various asset characteristics to the elasticity of demand for those assets and observe a decreasing trend in aggregate price impact multiplier over time similar to the one in this study, but slower in pace. In contrast, Haddad et al. (2021) find that the recent rise of passive benchmarks in recent years led to substantially more inelastic aggregate demand curves for individual stocks compared to the counterfactual scenario with perfect competition. Fuchs et al. (2023) discusses the theoretical implications of heterogeneity in asset-specific substitution and demand complementarities. This study contributes to this literature by documenting: (i) the heterogeneity among stocks in terms of substitutability, (ii) the primary role of substitutability in determining stocks' elasticity, and (iii) for the sample of large-cap stocks examined here, the diminishing price impact of S&P 500 index funds' demand shifts on all index firms over time.

To fully grasp the contribution of this paper in the shadow of existing literature, one must recognize that arbitrage is a contextual action. For the arbitrageur to perform their role–feeding the demand shifts and preventing large price impacts of those demand shifts and hence, reducing or eliminating mispricing–they must know why there is a demand shift and must be aware that it is unrelated to news or fundamentals. This study demonstrates that the ease of arbitrage due to improved substitution has enabled arbitrageurs to be well-prepared for the demand shifts generated by index funds, both in the context of additions and deletions, as well as during residual rebalancing of index incumbents. This, in turn, has reduced the sensitivity of prices to demand shifts-in other words, flattened the demand curves. However, it does not claim that the same flattening has occurred across the entire body of stocks or in all contexts. For example, in the case of mutual fund fire sales, when arbitrageurs, like everyone else, do not know why there is selling pressure, the price impact is large, but it eventually reverts when the market learns about the fund experiencing a fire sale (Honkanen and Schmidt, 2022).

The paper proceeds as follows. Section 2 sheds light on the demand shift measure construction and presents the theoretical framework, identification strategy, and hypotheses. Section 3 describes the data. Section 4 presents the empirical results, and the last section concludes.

1.2 Theoretical Framework and Identification

The first part of this section employs the weighing scheme utilized by S&P 500 to quantify the demand shifts that arise from the mechanical portfolio rebalancing of index funds. The second part outlines the identification strategy and formalizes the hypotheses to test.

1.2.1 Theoretical Framework

The weight of an index constituent j in S&P 500 portfolio at the closing of the trading day t is calculated as

$$w_t^j = \frac{P_t^j S_t^j IWF_t^j}{\sum_{i=1}^{N} P_t^i S_t^i IWF_t^i}, \tag{1.1}$$

where, P_t^j is the stock price, S_t^j is the total shares outstanding, and IWF_t^j is the Investable Weight Factor[3], all measured at the close of markets. N is the number of constituents in the index, typically 500 for S&P 500. IWF is the measure of float adjustment, which

[3] IWF of a stock is simply the ratio of floating shares to total outstanding shares. For instance, an IWF of 0.8 for stock j indicates that 80 % of total shares outstanding of that stock are freely tradable and available to the marketplace, and the rest 20% are held by strategic investors that are not expected to liquidate their position any soon. IWF's are crucial data used in determining constituents' weights. They are the key missing point preventing investors without S&P data subscription from replicating the index even if they have the list of included stocks.

is an attempt to enhance the investibility of the index by excluding shares of strategic shareholders in calculating firms' market capitalization[4].

Index level[5] is calculated as

$$Ind_t = \frac{\sum_{i=1}^{N} P_t^i S_t^i IWF_t^i}{Divisor_t}, \tag{1.2}$$

in which the denominator is the index divisor at the close of day t. The index divisor serves two purposes: first, dividing the free-float market value of the index by this factor does a scaling that helps market participants to work with a more easily handled number (e.g., 2000) rather than dealing with ten or more digits when reported in dollars. Second, and more crucial to this research, it is used as a level corrector to maintain the continuity of the index level following the implementation of corporate decisions, index reconstitutions, or other non-market-driven actions. Hence, the index divisor serves as the channel through which the difference in weight of added and deleted stocks affects the weights of other constituents in the index, ensuring that the sum of weights remains equal to 100%.

Equation (1.2) yields $\sum_{i=1}^{N} P_t^i S_t^i IWF_t^i = Ind_t * Divisor_t$, which enables us to rewrite equation (1.1) as

$$w_t^j = \frac{P_t^j S_t^j IWF_t^j}{Ind_t * Divisor_t}. \tag{1.3}$$

Therefore, the ratio of the weights of an index constituent j in the closing of two consecutive trading days would be

$$\frac{w_{t+1}^j}{w_t^j} = \frac{R_{t+1}^j}{R_{t+1}^{Ind}} * \frac{S_{t+1}^j}{S_t^j} * \frac{IWF_{t+1}^j}{IWF_t^j} * \frac{Divisor_t}{Divisor_{t+1}}, \tag{1.4}$$

[4]The list of shareholders that S&P deems strategic goes long. The most important ones are control groups, seat-holders, publicly traded companies, and government agencies. For an extensive list, please consult S&P Float Adjustment Methodology document at the S&P Dow Jones Indices web page https://www.spglobal.com/spdji/en/documents/index-policies/methodology-sp-float-adjustment.pdf

[5]In this paper, the term "index level" consistently pertains to the S&P 500 price index level, the same metric that is primarily used in news and statistical reporting. The S&P 500 total and net return indices are derived from the price index, incorporating necessary adjustments related to dividends and taxes.

where R_t^j and R_t^{Ind} are respectively the gross returns of the stock j and the S&P 500 price index in day t. The decomposition in Equation (1.4) shows that any change in the weight of a stock from one day to the next comes from one or some of these four ratios.

The first ratio in Equation (1.4) represents the stock price growth relative to the index. It is important to note that for an ideal index fund that closely replicates the index portfolio, changes in this component of weight do not necessitate portfolio rebalancing[6]. In other words, if there are no shocks affects the portfolio composition and only price movements occur, the value-weighted portfolio will remain correctly value-weighted.

In contrast to the first ratio, index funds must rebalance their portfolios if any of the remaining three ratios in Equation 1.4 is not equal to one. Although daily changes in index divisor are widespread, the amount of weight change that firms incur only because of the modification in the index divisor is free of information about each specific firm. Even the changes on the IWF and outstanding shares do not contain new information about stocks since index maintenance requires a holding period before implementing the changes on their calculations[7]. Except for rare cases, these numbers are often revised during annual or quarterly rebalancings.

Aiming to find a measure of fund flows implied by the index funds' mechanical rebalancings, I define the surprises to the weights of index incumbents that require rebalancing for index followers as Δw_{t+1}^j that is calculated as follows:

$$\Delta w_{t+1}^j = w_t^j * \frac{S_{t+1}^j}{S_t^j} * \frac{IWF_{t+1}^j}{IWF_t^j} * \frac{Divisor_t}{Divisor_{t+1}} - w_t^j. \tag{1.5}$$

[6]For instance, assume the index return is zero in a day t ($R_t^{Ind} = 1$) and some stock j has a return of 1% on that day ($R_t^j = 1.01$). Also, assume the other three ratios are equal to one, which means there is no stock repurchase or equity issue for this stock, no change in the number of floating stocks, and no change in the index divisor on that day. Suppose the index fund already had the right number of shares at the previous day's closing (i.e., proportional to the weight of this stock in the index at the close of the market). In that case, it will automatically have the correct number of shares at the closing of day t, which is, in fact, the same number of shares! However, this number of shares results in 1% more weight in the index today than the previous day, precisely as it should. I assumed this fund had no net inflows or outflows on this day.

[7]For this reason, I also use SPDJI data as reference for the number of shares outstanding in the calculation of weight surprises.

Intuitively, Δw^j_{t+1}, is the weight of stock j in the index portfolio one moment after the implementation and one moment before it, both of which are out of the trading session[8]. In the empirical analysis, I drop a few observations that went through stock splits or issued stock dividends on the event days, as these are corporate action with separate effects of prices and number of shares with no overall effect on market capitalization. Section A.1 in the Appendix explains the dynamics of the weight shifts with a simple example.

I define the measure of fund flows implied by mechanical (information-free) rebalancing as follows:

$$MRF^j_{t+1} = \frac{A_{t+1}\Delta w^j_{t+1}}{S^j_t P^j_t}, \tag{1.6}$$

in which MRF stands for Mechanical Rebalancing Flow, and $A_{t+1} = \sum A^F_{t+1}$ is the total amount of dollars invested in passive funds that are benchmarked against S&P 500 in the opening of the day $t+1$. MRF measures the surprises in the inelastic demand of an ideal index fund, similar to the measure constructed and used in Pandolfi and Williams (2019) for a weight-capped bond market index. Here, instead of the cap on the constituents' weights, MRF is built upon the weight spillovers on index incumbents through the index divisor changes. MRF also has similarities with the measure of benchmarking intensity in Pavlova and Sikorskaya (2022) and can be motivated by a theory like theirs.

To put it simply, MRF^j_t captures the amount of money that will flow into or out of an stock j at day t, relative to its market value in the previous trading day, as a result of mechanical rebalancing of index followers regardless of the stock's individual characteristics or fundamentals. Based on this intuitive definition, I can extend the definition of MRF to additions and deletions, in which case I will use w of the added stock or $-w$ of the deleted stock instead of Δw in the MRF formula. Note that, for example, for an addition in day t, Δw reduces to w simply because that stock had zero weight in the index portfolio on the previous day.

[8]Implementation of stock changes happen after the close of trading on the day prior to effective day so that on the effective day when the market opens, the added stock is already in the index portfolio and the removed one is no longer in it.

1.2.2 Identification and Hypotheses

The identification assumption in pinning down the causal effect of demand in this paper hinges on the fact that a stock, witnessing an exogenous increase (decrease) in its weight by an amount Δw_t^j, and a stock newly added to (removed from) the index on day t with a weight of $w_t^i = \Delta w_t^j$, will undergo a similar demand shock emanating from index funds, commensurate with their respective market capitalizations. Consequently, any differences in the abnormal returns observed for these stocks must be attributed to factors unrelated to the mechanical demand shifts of index funds. Following the classification in Afego (2017), I refer to these factors collectively as the informational components of the index effect.

The central hypothesis focuses on testing the effect of MRF as a measure of information-free demand shifts on prices and other variables of interest. Formally, I will run the following regression in time-invariant settings:

$$y_{jt} = \beta MRF_t^j + \phi X_{jt} + \theta_t + \theta_j + \varepsilon_{jt}, \tag{1.7}$$

where y_{jt} is the dependent variable, θ_t and θ_j are fixed effect dummies, and X_{jt} represents other control variables. Notably, β will quantify the causal effect of non-fundamental demand shifts on the outcome variables. To study the temporal evolution of effects, I will employ the model with time-varying sensitivity, estimating the conditional model by replacing β with β_t in the equation as follows:

$$y_{jt} = \beta_t MRF_t^j + \phi X_{jt} + \theta_t + \theta_j + \varepsilon_{jt}, \tag{1.8}$$

where $\beta_t = \sum_p \beta_p \mathbb{1}(period = p)$. This approach enables me to estimate different coefficients for different periods by dividing the sample period into multiple intervals. For the price impact regressions, I use two-year non-overlapping periods,[9] while for other

[9] I also attempt to use yearly dummy variables. The results exhibit a similar overall pattern but are

results, I utilize decades or simply pre- and post-2000 periods, as these are sufficient for interpretation.

1.3 Data

The index composition changes are sourced from Siblis Research. The primary dataset for this study consists of the list of S&P 500 constituents and their daily weights in the index. For the post-2000 period, I obtain this data directly from S&P Dow Jones Indices-SPDJI. For the pre-2000 period, I manually construct the daily portfolios based on the portfolio changes in Siblis Research data and the first available direct observation from SPDJI[10]. Daily stock data and quarterly fundamentals are retrieved from CRSP and Compustat. Risk-free returns and market returns are sourced from Kenneth R. French's website[11]. Data on index funds, including their prospectus benchmarks and assets under management, are from Morningstar. The sample period spans from January 1990 to June 2021, based on the availability of daily index constituents' data from SPDJI[12]. All return and turnover variables from CRSP that are winsorized at the 0.5% level at both tails.

The final sample contains 745 additions and 740 deletions which occurred in 629 distinct effective dates that were later used to make the sample of index incumbents. I

noisier, as some years, such as 2003, have significantly fewer events than others. Employing two-year periods smooths out these idiosyncrasies. Additionally, using three-year dummy variables yields results comparable to those from the two-year approach.

[10]This approximation of portfolio weights is particularly appropriate given that S&P transitioned to using the floating number of shares and employed investable weight factors (IWFs) only in 2005. This implies that all IWFs prior to 2000 are equal to one. In calculating the MRF for the pre-2000 period, the only unknown is the number of shares utilized by S&P, which may slightly differ from the values in CRSP. To minimize measurement error, I assume that the ratios of shares outstanding on the effective date and the day prior to it are always equal to one, as I only require the ratio and not the individual actual values. I also attempted to use the CRSP number of shares; however, in Equation 1.5, the changes in the divisor are small, making the measurement error in changes of shares outstanding too large to reasonably detect the effect. Additionally, I verified this in the post-2000 period, where I can compare the number of shares in SPDJI data and CRSP. It is evident that there are many small changes in the CRSP number of shares that S&P does not necessarily accommodate, possibly to avoid unnecessary weight adjustments.

[11]www.mba.tuck.dartmouth.edu/pages/faculty/ken.french/index.html

[12]Start of the sample period is chosen so because before 1990 S&P was not announcing the portfolio changes in advance.

further divide the index events into forced and discretionary based on the trading status of the deleted firm after the event. Forced index events are those in which the index deletion was delisted entirely from the exchange within two weeks[13] from the deletion effective date. Regardless of why these stocks were delisted, the index committee had to drop them from the index simply because they wouldn't be trading anymore. The discretionary deletions are those that were not forced. In other words, the index committee chose to remove discretionary deletions, even though they continued trading.

Table 1.1 presents the summary description of the data, highlighting three relevant facts. First, regarding average market capitalization, index additions are almost twice larger than discretionary index deletions but smaller than forced index deletions. This discrepancy in size and hence portfolio weights is the exogenous source of variation which allows pinning down the causal effect of demand shifts on prices of index incumbents. Second, index additions exhibit a smaller bid-ask spread, while their turnover is comparable to that of deletions. Third, additions have lower market betas and are less volatile than discretionary deletions but this reverts when compared to forced deletions. Overall, it seems like additions are larger, less risky (both in terms of beta and range), and more liquid (smaller bid-ask spread) in comparison to discretionary deletions. Furthermore, the distance between announcement day and effective day for both additions and deletions is highly concentrated around five trading days. For index incumbents, on each reconstitution day, I take the maximum of this distance between all additions and deletions on that day.

Data on index funds were obtained from Morningstar. From the universe of passive funds, I collected data for those tagged as index funds (including both open-end funds and ETFs) that designated the S&P 500 as their primary prospectus benchmark. I also excluded a few fund-of-funds (typically international index funds that achieve this by

[13] I verify that the choice of the period does not have any meaningful effect on the results. From all the deletions that were delisted within three months of the effective date, more than 89% were already delisted one day after the effective day.

Table 1.1: Summary Statistics

Variable	Mean	SD	p25	p50	p75	N
Index Incumbents						
Market Cap (B$)	23.99	52.41	4.44	9.76	21.96	308753
$Weight_{Eff}(\%)$	0.20	0.35	0.05	0.09	0.19	308753
MRF (*10^4)	-2.52	10.06	-2.42	-0.17	0.02	308753
β^{CAPM}	0.99	0.45	0.70	0.94	1.22	308753
IWF	0.98	0.07	1.00	1.00	1.00	308753
$AvgTO_M(\%)$	0.83	0.88	0.35	0.59	1.00	308753
$AvgBidAsk_M(\%)$	0.48	0.72	0.03	0.10	0.74	289022
$AvgRange_M(\%)$	2.86	1.84	1.77	2.36	3.32	308753
Ann to Eff (days)	5.40	4.72	3.00	5.00	6.00	308753
Index Additions						
Market Cap (B$)	11.92	25.77	5.09	7.94	12.48	745
$Weight_{Eff}(\%)$	0.10	0.13	0.05	0.07	0.10	744
MRF (*10^4)	330.16	211.87	203.53	291.63	433.07	671
β^{CAPM}	1.10	0.57	0.72	1.02	1.38	652
IWF	0.96	0.11	1.00	1.00	1.00	744
$AvgTO_M(\%)$	1.10	1.08	0.43	0.77	1.34	661
$AvgBidAsk_M(\%)$	0.40	0.58	0.03	0.11	0.65	631
$AvgRange_M(\%)$	3.31	1.89	2.09	2.76	3.88	661
Ann to Eff (days)	5.69	4.70	4.00	5.00	6.00	745
Discretionary Index Deletions						
Market Cap (B$)	2.58	3.66	0.50	1.64	3.41	326
$Weight_{Eff-1}(\%)$	0.03	0.08	0.01	0.01	0.02	326
MRF (*10^4)	-311.95	183.04	-441.82	-298.66	-166.99	304
β^{CAPM}	1.18	0.62	0.75	1.13	1.51	326
IWF	0.97	0.09	1.00	1.00	1.00	326
$AvgTO_M(\%)$	1.73	1.83	0.54	1.17	2.16	326
$AvgBidAsk_M(\%)$	0.93	1.44	0.06	0.21	1.29	298
$AvgRange_M(\%)$	4.55	3.56	2.31	3.42	5.29	326
Ann to Eff (days)	5.55	2.45	4.00	6.00	6.00	325
Forced Index Deletions						
Market Cap (B$)	12.68	14.04	4.02	7.74	15.56	330
$Weight_{Eff-1}(\%)$	0.12	0.13	0.04	0.08	0.15	398
MRF (*10^4)	-284.60	145.20	-359.66	-270.49	-191.91	330
β^{CAPM}	0.87	0.53	0.51	0.78	1.12	413
IWF	0.98	0.07	1.00	1.00	1.00	412
$AvgTO_M(\%)$	1.04	0.94	0.49	0.77	1.24	414
$AvgBidAsk_M(\%)$	0.54	1.02	0.03	0.11	0.74	392
$AvgRange_M(\%)$	2.18	2.16	0.80	1.60	2.74	414
Ann to Eff (days)	5.49	5.51	3.00	5.00	6.00	414

This table presents a summary description of the sample stocks. Deleted firms are categorized into forced and discretionary deletions. Forced deletions refer to firms that are completely delisted from their exchange within two weeks following their removal from the index. IWF denotes the proportion of floating shares relative to total outstanding shares. The reported β is derived from the market model, utilizing an estimation window of one year that concludes 20 days prior to the effective date, requiring a minimum of 100 observations within this window. The variable IWF represents the investable weight factor. $AvgTO_M(\%)$ indicates the average turnover (volume of trades divided by the number of shares). $AvgBidAsk_M(\%)$ is the average bid-ask spread (expressed as a percentage of the mid-price). $AvgRange_M(\%)$ refers to the average price range (high price minus low price, as a percentage of the previous day's closing price and adjusted for stock splits) serving as a measure of volatility. The averages for $AvgTO_M(\%)$, $AvgBidAsk_M(\%)$, and $AvgRange_M(\%)$ are calculated over a one-month window ending 20 days prior to the effective dates and requiring a minimum of 9 valid observations. The table also shows the number of days after announcement in which the changes are made effective. If applicable, the unit of measure is indicated before the variable's name. The sample period is (1990-01, 2021-06).

purchasing U.S.-based index ETFs) from the remaining sample[14] and leveraged funds. When available, I used the daily assets under management (AUM) labeled as “Fund Size - Comprehensive (Daily)." If daily data were unavailable, I relied on the monthly values in the variable “Total Market Value (Long)" and interpolated interim values, assuming that inflows and outflows occurred at the end of each month[15][16] This procedure resulted in a total of 200 index funds in my sample, although this figure can be misleading due to the highly skewed distribution of fund sizes. The four largest players in the indexing industry-Fidelity, Vanguard, BlackRock (including iShares), and State Street-together account for an average of 79% of the total S&P 500 passive AUM over my sample period, with their share exceeding 85% during certain intervals. It is noteworthy that, despite the rigor of the methodology employed, the resulting numbers are similar to alternative approaches. Greenwood and Sammon (2024) report broadly similar estimates of AUM for S&P 500 index funds using a fundamentally different method based on portfolio holdings.

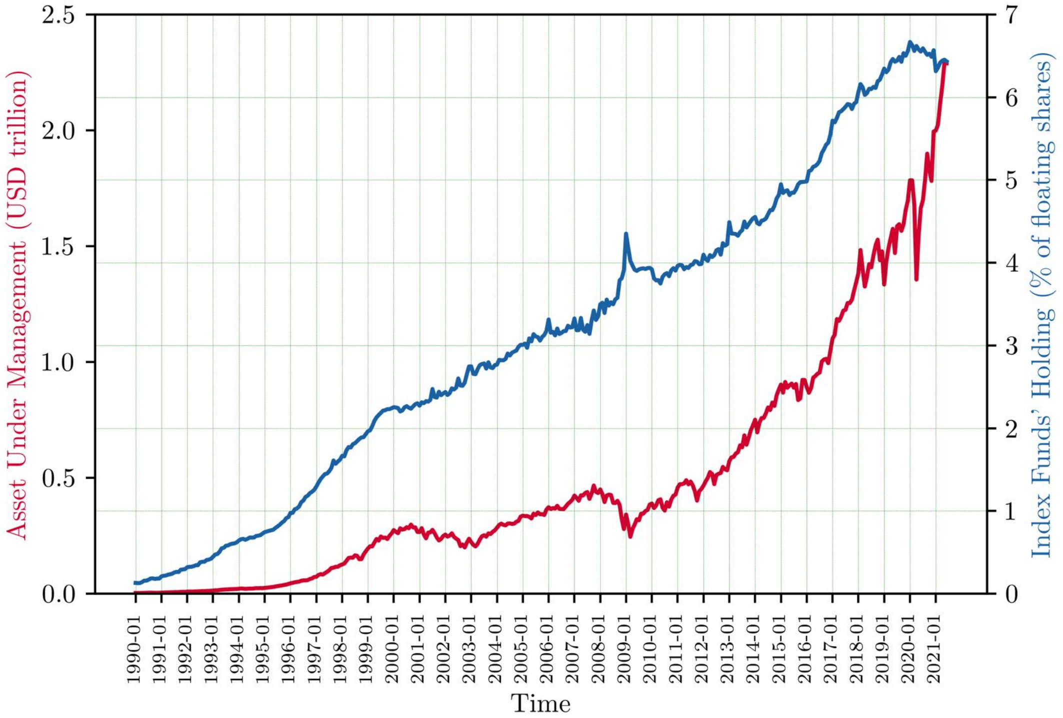
Figure 1.1 shows the total assets under management of index funds and their aggregate percentage holding of an exemplary index firm. In three decades, the total AUM of index funds rose from almost zero to USD 2.5 trillion, and they were holding an average of about 7% of total shares outstanding of index firms. The growth of the percentage holding also shows that most of the rise in the AUM of index funds comes from money inflow to these funds as opposed to organic capital gains in the stock market, since without the inflows the percentage would remain relatively constant.

[14]I retained 62 international index funds that are not classified as fund-of-funds, which account for only 1.2% of the total assets under management (AUM) of index funds on average over the sample period. All results remain robust even after excluding these funds.

[15]This approach is equivalent to assuming that the daily returns of the index funds and the index are equal. For robustness, I also examined an alternative assumption for funds with missing daily AUMs, where I assumed AUM increased linearly between two end-of-month observations. This alternative assumption did not affect the results, either qualitatively or quantitatively.

[16]For one index fund from Vanguard (FundId FSUSA002QH), the daily and monthly time series were not compatible. I checked their website and verified that the monthly time series contained the correct values, and I used that data.

Figure 1.1: Assets Under Management and Percentage Ownership of Index Funds



This figure illustrates the total assets under management of passive index funds benchmarked to the S&P 500, along with their aggregate percentage ownership of index stocks. Refer to the main text for details on the selection mechanism of index funds. The red line represents the assets under management in USD trillions, while the blue line indicates the average percentage of an index firm held collectively by index funds. The sample period is (1990-01, 2021-06).

1.4 Empirical Results

In the initial part of this section, I examine the effect of index reconstitutions on index incumbents. I begin by analyzing the aggregate price impact over full sample, followed by a temporal evolution analysis. Subsequently, I explore how arbitrage risk contributes in explaining the heterogeneity of price impact experienced in response to demand shifts. Finally, I investigate how changes in demand affect trading volumes surrounding index reconstitutions. In the second subsection, I will assess the index effect associated with index additions and deletions, aiming to quantify the extent to which demand shifts alone can explain these observed outcomes.

1.4.1 Index Incumbents

This sub-section demonstrates that the defined measure of mechanical demand, MRF, explains the price reaction for index incumbents, quantifies the price impact multiplier both in the full sample and over time, and establishes models to create a counterfactual information-free demand response for additions and deletions in the subsequent part of empirical results.

Price Impact

Table 1.2 presents the association of MRF with returns using the regressions in Equation (1.7). Each regression studies the impact of demand shift on a specific measure of price changes. Consistent with downward sloping demand curves hypotheses, all else equal, price reactions are larger for stocks experiencing larger shifts in their weights. I focus first on the full sample estimates and then divide the sample into three parts (each almost a decade) to observe the structural breaks. All regressions include controls in addition to stock and year fix effects. Control variables include log MV (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All abnormal returns in the paper are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations.

The full sample estimates presented in Table 1.2 reveal that a MRF of 1%, equivalent to purchasing 1% of shares outstanding, results in an approximately 51 bps increase in price on the effective day and a 78 bps price increase on the day preceding the effective day. Of the total return realized on the effective day, approximately half occurs during the

Table 1.2: MRF and Price Impact

Dependent Variable	(1) Ret_{Eff}	(2) Ret_{Eff-1}	(3) Ret^{open}_{Eff}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff-2}_{Ann}	(6) CAR^{Eff}_{Eff-1}	(7) AR_{Eff}	(8) AR_{Eff-1}
				Full Sample (1990-2021)				
MRF	0.514***	0.780***	0.235***	0.411***	0.188**	0.224***	0.137**	0.088*
	(7.07)	(8.76)	(5.94)	(3.57)	(2.28)	(2.63)	(2.33)	(1.83)
Adj. R-sq	0.016	0.026	0.017	0.006	0.005	0.002	0.002	0.003
N	289020	289020	288158	288997	288997	288997	289020	288997
				Early Sample (1990-1999)				
MRF	0.869***	1.150***	0.254***	0.348**	0.175	0.174**	0.059	0.116**
	(14.23)	(18.75)	(10.05)	(2.50)	(1.57)	(2.22)	(0.98)	(2.05)
Adj. R-sq	0.010	0.017	0.007	0.017	0.017	0.004	0.005	0.003
N	80042	80042	79189	80033	80033	80033	80042	80033
				Mid Sample (2000-2010)				
MRF	-0.167	0.733***	0.095	1.013***	0.426***	0.587**	0.286	0.301**
	(-1.04)	(2.60)	(0.79)	(3.16)	(2.73)	(2.20)	(1.61)	(2.46)
Adj. R-sq	0.017	0.031	0.018	0.009	0.007	0.003	0.003	0.005
N	118685	118685	118676	118676	118676	118676	118685	118676
				Late Sample (2011-2021)				
MRF	0.315*	-0.118	0.331***	-0.053	-0.059	0.006	0.197	-0.191
	(1.71)	(-1.01)	(3.41)	(-0.27)	(-0.40)	(0.03)	(1.63)	(-1.59)
Adj. R-sq	0.020	0.007	0.022	0.010	0.007	0.005	0.003	0.008
N	90290	90290	90290	90286	90286	90286	90290	90286
				Difference: Mid - Early				
	-1.036***	-0.417	-0.159	0.665*	0.251	0.413	0.227	0.185
	(-6.03)	(-1.45)	(-1.29)	(1.90)	(1.31)	(1.49)	(1.21)	(1.37)
				Difference: Late - Mid				
	0.482**	-0.851***	0.236	-1.066***	-0.485**	-0.581*	-0.089	-0.492***
	(1.97)	(-2.79)	(1.53)	(-2.84)	(-2.26)	(-1.74)	(-0.41)	(-2.87)

This table presents the results of estimating the panel regression specified in Equation (1.7) for index incumbents. The full sample comprises all incumbent observations on reconstitution days throughout the entire sample period (1990-01 to 2021-06). Results are then replicated separately for three sub-samples: Early (1990-1999), Mid (2000-2010), and Late (2011-2021). The bottom rows assess the statistical significance of the differences between sub-sample estimates. The independent variable is MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. All abnormal returns are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations. T-statistics based on standard errors clustered by stocks are reported in parentheses. Significance levels are denoted as: $^*p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

opening[17] batch auction[18]. The primary outcome variable in Table 1.2 is the cumulative abnormal return of stocks in response to demand shifts over the entire period from the announcement to the effective day, as presented in Column (4). This indicates that, in response to a 1% demand shift, the price of stocks, on average, would increase by 41 bps, with most of this increase occurring on the effective day and the preceding day, as detailed in the final four columns. Furthermore, Columns (4) to (8) that focus on abnormal returns emphasize that the observed heterogeneity in the price movements of stocks is attributable to the varying surprises in their portfolio weights which cannot be fully explained by market-wide risk factors.

As I will show later in Section 1.4.2, the price reactions to index additions and deletions have been falling through the years, despite the sharp increase in passive investing that was particularly documented for S&P 500 in Figure 1.1. The diminishing magnitude of the index effect had been reported in other recent studies, for instance, Bennett et al. (2023); Greenwood and Sammon (2024); Patel and Welch (2017) among others. improved market efficiency stands as the most reasonable explanation for this stark and somewhat surprising observation, but it is not easy to test this hypothesis using the data of index additions and deletions themselves. This is because one cannot disentangle easily between the decreasing price impact related to the demand-driven component of the index effect or the less informative recent index decisions associated with the non-demand-driven part of it. Furthermore, the recent increase in switching between S&P500 stocks and their smaller counterparts, S&P 400 mid-cap and S&P 600 small-cap indices, makes identifying the possible scenarios even harder (Vijh and Wang, 2022; Kumar et al., 2023). The

[17] I define the daily opening return of stocks as $(P^{open}t/P^{close}t-1-1)$, adjusted for dividends and stock splits.

[18] The opening batch auction is a process that occurs prior to the commencement of continuous trading for the day, aimed at establishing the opening price for securities. During this interval, buy and sell orders are matched and executed at a single price point. In contrast, the closing auction serves a similar function, determining the closing price for securities by matching buy and sell orders at a single price point. Market participants can submit orders to buy or sell securities at the closing price during this auction, typically conducted a few minutes before the trading day concludes. The closing price is determined at the price point where the maximum volume of shares can be transacted.

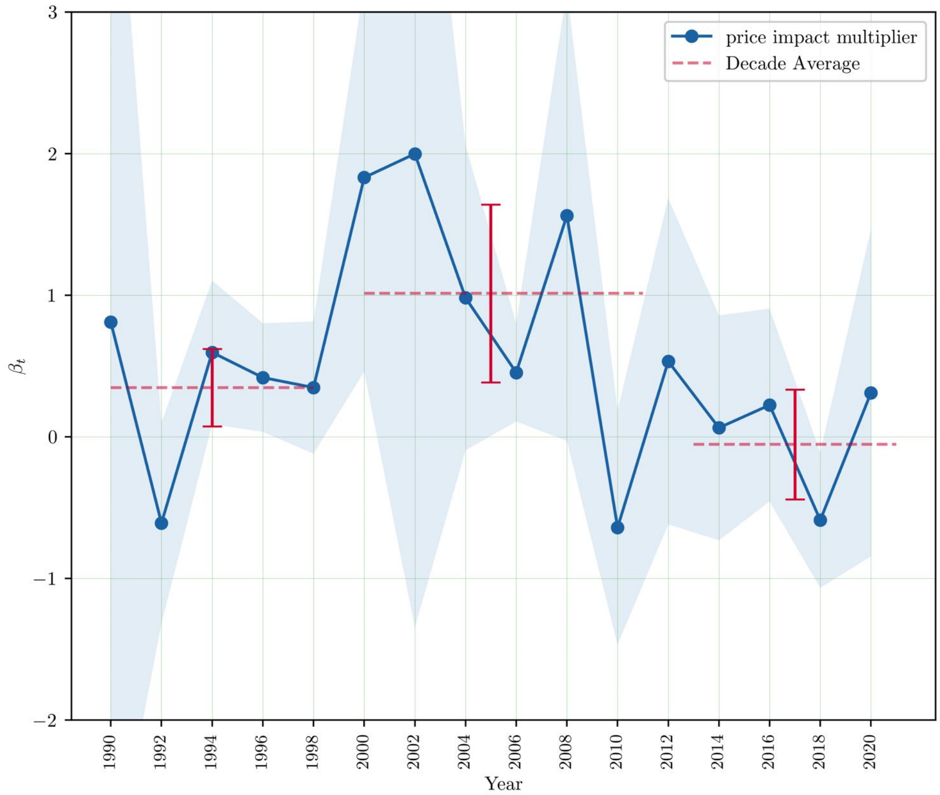
primary contribution of this paper is to solve this problem by the novel identification that moves the focus from demand shift on additions and deletions to an information-free demand shift on index incumbents.

The subsequent rows of Table 1.2 replicate the full sample results separately for the three decades encompassed in the sample[19]. A broad observation from this analysis is that the results exhibit striking differences across the three sub-samples. In the early period (1990-1999), the coefficients are substantial, with most demonstrating significantly positive values, indicating downward-sloping demand curves. In the mid-sample period (2000-2010), the coefficients are even larger and more significant, suggesting an increase in the general slope of the demand curves. Conversely, the coefficients for the late part of the sample are predominantly close to zero and statistically insignificant, reflecting a notable flattening of the stocks' demand curves. These findings, which motivate a time-varying analysis in the remainder of this section, are further corroborated in the last two rows, where I verify that the estimated price impact multipliers differ significantly across the three sub-periods.

To provide a more precise analysis of the evolution of the price impact multiplier, I employ the dynamic regression framework outlined in Equation (1.8), incorporating a time-varying coefficient to examine the temporal trend of price reaction sensitivity to index funds' demand shifts. I utilize the CAR_{Ann}^{Eff} as the measure of price reaction to index decisions, ensuring that the outcome variable encompasses both the anticipated and unanticipated components of the price reaction. This choice is further justified by the intention to use the results of this regression in the subsequent section to construct a counterfactual scenario regarding the index effect on additions and deletions. Such a selection of the time window facilitates an apples-to-apples comparison between the reality and counterfactual scenarios studied.

[19]The sample spans 31.5 years. To avoid categorizing it into four decades, I include the year 2010 in the second decade and the first six months of 2021 in the last decade. Consequently, the three sub-samples comprise 10, 11, and 10.5 years, respectively. No results would materially change with any alternative division of the sample.

Figure 1.2: Price Impact Across Time



This figure shows the estimates of price impact over time where the estimates come from the panel regression specified in Equation (1.8), in which the outcome variable is the CAR_{Ann}^{Eff}, total cumulative abnormal return of incumbents from announcement to effective day of index changes, inclusive of both dates. β_t estimates using dummy variables from a division of sample time into two-year periods (1990-01, 2021-06)-sixteen periods in total (the last period is one and half years). Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. Standard errors clustered by stocks and 95% confidence intervals are shown in light shadow around the means. The estimates for three sub-periods, Early, Mid, and Late from 1.2 are also shown in the graph.

Figure 1.2 illustrates the price impact estimates of index funds derived from the time-varying model. The graph reveals that the average price impact of index funds' demand shifts initially increased throughout the 1990s, peaking around 2000, and subsequently declined over the past two decades. The sensitivity of prices to demand shifts has decreased from approximately 2 at the turn of the century to nearly zero in the later period. This direct evidence of an overall enhancement in market efficiency suggests that market participants, including arbitrageurs and market makers, have significantly improved their ability to provide liquidity to index funds. Furthermore, the figure indicates that the aggregate price impact is closely linked to the market's overall capacity to accommodate demand shifts. For example, during periods of financial crises and the COVID-19 pandemic, when market conditions and arbitrage capital were constrained, the price impact multiplier surged dramatically. The following section elaborates on how the ability of arbitrageurs to absorb demand shifts critically influences the price impact multiplier and, consequently, the slope of demand curves.

Arbitrage Risk

Gromb and Vayanos (2010) utilize a two-stock model to demonstrate that the price impact of a demand shock is influenced by several factors: (i) the magnitude of the shock itself, as arbitrageurs require greater compensation for bearing higher risk; (ii) the unconditional variance of the asset affected by the shock, where increased volatility necessitates higher compensation from arbitrageurs; and (iii) $(1-\rho^2)$ where ρ is the correlation of the asset with its substitutes (specifically the only other stock in their model), as lower correlation constrains arbitrageurs' ability to hedge effectively. Consequently, the degree of substitutability of stocks with their close substitutes is crucial in explaining the abnormal returns observed across stocks in response to certain shift in demand. Please see Appendix A.3 for a more detailed discussion. The product of the latter two determinants of price impact in their model, unconditional variance and $(1-\rho^2)$, is exactly the residual variation in the stock prices that an arbitrageur cannot hedge using available substitutes, and is called arbitrage risk in Wurgler and Zhuravskaya (2002). In this section, I first establish that arbitrage risk serves as a channel that moderates the effect of demand shifts on prices, thereby providing an empirical test for their theoretical predictions. Subsequently, I analyze how the overall trend of this variable across stocks has evolved over time and its relationship to the flattening of demand curves and the price impact estimates presented

Table 1.3: Arbitrage Risk Summary Description

Variable	Mean	SD	p25	p50	p75	N
Index Incumbents						
A_1 (*10^4)	4.11	5.67	1.35	2.44	4.73	308753
A_2 (*10^4)	3.87	5.45	1.26	2.32	4.43	305430
A_0 (*10^4)	5.87	8.38	1.98	3.46	6.52	308753
E_1 (*10^4)	1.76	3.47	0.36	0.75	1.62	308753
E_2 (*10^4)	2.02	4.18	0.37	0.82	1.96	305430
R_1	0.28	0.18	0.13	0.25	0.40	308753
R_2	0.32	0.22	0.13	0.28	0.47	305430

This table provides an overview of arbitrage risk and related metrics for the incumbent stocks. Two stock-specific arbitrage-risk measures, following the methodology of Wurgler and Zhuravskaya (2002), were constructed using daily returns over a calendar window from $[-365, -20]$ days. The first measure, A_{1i}, represents the variance of residuals from the model $R_{it} - R_{ft} = \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it}$, where R_{mt} denotes the return on CRSP's value-weighted index and R_{ft} is the risk-free rate. The second measure, A_{2i}, measures the residual variance of $R_{it} - R_{ft} = \beta_{1i}(R_{\text{sub1it}} - R_{ft}) + \beta_{2i}(R_{\text{sub2it}} - R_{ft}) + \beta_{3i}(R_{\text{sub3it}} - R_{ft}) + \varepsilon_{it}$, with $R_{\text{sub1it}}, R_{\text{sub2it}}$, and R_{sub3it} representing returns on three closest industry-, size-, and book-to-market-matched "substitute" stocks, matched using quintiles of size and book-to-market and utilizing Fama and French 49 industry classifications (Fama and French, 1997). See Wurgler and Zhuravskaya (2002) for more details. Total variance of excess returns $A_{0it} = \text{Var}(R_{it} - R_{ft})$ is estimated over the same interval as the arbitrage-risk measures. E_{1i} represents the "explained" variance $\text{Var}(R_{it} - R_{fi}) - A_{1i}$, while R_{1i} is the ratio of explained variance E_{1i} to total variance A_{0i}. E_{2i} and R_{2i} are defined analogously. All variables except ratios in the table are multiplied by (10^4) to ease reading. A more detailed version of this table including additions and deletions is also available in Appendix section A.4 The sample period spans from January 1990 to June 2021.

in the previous section.

In measuring arbitrage risk, I adopt the methodology outlined in Wurgler and Zhuravskaya (2002), which utilizes the variance of the residuals from regressing the excess returns of a stock against the excess returns of its substitutes over the calendar period [-365, -20][20]. This approach ensures zero net investment of arbitrageurs since it measures the residual variance when they are long one dollar on the stocks under study and short a net one dollar on substitutes, including the risk-free rates. Essentially, arbitrage risk quantifies the portion of total stock price variation that cannot be hedged using available substitutes. Therefore, if a stock has a perfect substitute, its arbitrage risk will be zero. Wurgler and Zhuravskaya (2002) define two measures of arbitrage risk: the first

[20]No constant is included in these regressions because a constant lacks any economic interpretation as a substitute. Therefore, it's not strictly accurate to decompose total variance into residual variance (arbitrage risk) and explained variance. As a result, occasional negative R-squared values may occur.

measure (A_1) considers only the market and the risk-free rate as the substitute for all stocks, while the second measure (A_2) uses three stocks matched by industry and closest in book-to-market ratio and size to the stock under study[21] plus the risk-free rate.

Table 1.3 outlines the estimation procedure and presents summary statistics of arbitrage risk measures in which I also compute summary statistics for the total variance of stocks' excess returns, including both the absolute amount and the fraction explained by substitutes' returns. The key takeaway from Table 1.3 is that while index incumbent stocks lack perfect substitutes, a significant portion of their price variation can be hedged using close substitutes. For the median stock in the S&P 500, approximately one third of the total variation can be mitigated through cross-stock arbitrage. A more detailed version of Table 1.3 that includes also additions and deletions is available in Appendix section A.4. When comparing these ratios to those from the original study by Wurgler and Zhuravskaya (2002), we find that the portion of total variation that can be hedged by arbitrageurs with only three substitutes is already three times larger than during their study period.

I add arbitrage risk to the previously studied price impact regressions to find out how the heterogeneity of price impact experienced by different incumbent relates to their arbitrage risk, or substitutability. To enhance the interpretability of the regression results, I standardize these variables by dividing them by their standard deviations. The underlying regression models to test are versions of the following equation

$$CAR^{Eff}_{Ann,jt} = \beta_0 MRF^j_t + \beta_1 MRF^j_t \cdot A^N_{k,jt} + \beta_2 A^N_{k,jt} + \phi X_{jt} + \theta_t + \theta_j + \varepsilon_{jt}, \qquad (1.9)$$

where $A^N_{k,jt}$ is either A_1 or A_2 (k = 1 or 2) of stock j at day t divided by the unconditional

[21]Similar to this approach of constraining the universe of potential substitutes, Pontiff (1996) selects ten open-end funds as the set of potential substitutes for each closed-end fund in his study. Li and Lin (2022) also categorize stocks based on size and book-to-market ratio, and then decompose the underlying demand into different levels of granularity.

Table 1.4: MRF and Arbitrage Risk

	(1) CAR_{Ann}^{Eff}	(2) CAR_{Ann}^{Eff}	(3) CAR_{Ann}^{Eff}	(4) CAR_{Ann}^{Eff}	(5) CAR_{Ann}^{Eff}	(6) CAR_{Ann}^{Eff}	(7) CAR_{Ann}^{Eff}
MRF	0.411***	0.210	0.346	0.071	0.146	0.283	0.099
	(3.57)	(1.51)	(1.55)	(0.46)	(0.95)	(1.20)	(0.61)
A_1^N		0.002***	0.006***	0.001***			
		(5.57)	(3.98)	(3.95)			
$MRF \times A_1^N$		0.256	-0.060	0.550***			
		(1.31)	(-0.19)	(3.34)			
A_2^N					0.002***	0.003*	0.001***
					(5.66)	(1.92)	(4.74)
$MRF \times A_2^N$					0.358*	0.040	0.604***
					(1.65)	(0.13)	(3.78)
Sample	All	All	Pre 2000	Post 2000	All	Pre 2000	Post 2000
N	288997	288997	80033	208963	286060	78208	207851
Adj. R-sq	0.006	0.006	0.019	0.008	0.007	0.018	0.008

This table reports the results of estimating panel regression in Equation (1.9) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include MRF_t^j, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio, and the arbitrage risk measures divided by their standard deviation for the ease of interpretation. Arbitrage risk measures are calculated as the variance of residuals from regressing one stock's excess return on the market's excess return over a window of [-365, -20] prior days (A_1), or on the excess return of the three closest industry-, size-, and book-to-market-matched "substitute" stocks (A_2). These substitutes are matched using quintiles of size and book-to-market, employing the 49 industry classifications from Fama and French (1997). For further details, refer to Wurgler and Zhuravskaya (2002). Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: *p < 0.10; **p < 0.05; ***p < 0.01.

standard deviation of this variable. The coefficients of interest are β_1 and β_2.

Table 1.4 illustrates the relationship between arbitrage risk and price impact. The dependent variable is always the total cumulative abnormal return of index incumbents during the period from the announcement to the effective day of index changes, inclusive of both dates. Regression (1) presents the baseline result extracted from Table 1.2 for comparison. Regressions (2) to (4) utilize A_1 as the measure of arbitrage risk, while regressions (5) to (7) employ A_2. In both cases, I first examine the full sample results and then focus separately on the pre- and post-2000 periods. The choice of 2000 is made because, as documented in the previous section, the flattening of demand curves and, in a sense, the regime change began broadly around that time. However, in untabulated results, I verify that the findings would be similar if the division year were adjusted to any year around 2000.

The most important finding of Table 1.4 is that arbitrage risk serves as the channel through which demand shocks impact prices in the following manner: In all regressions, regardless of the arbitrage risk measure used or the sample period considered, once the interaction between MRF and arbitrage risk is included in the baseline regression, the coefficient of MRF is no longer significantly different from zero. This implies that if a stock has perfect substitutes (and thus zero arbitrage risk), demand shifts will not affect its price. In other words, exactly in line with the prediction in theory (Vayanos and Woolley, 2013), demand shifts influence prices only in the presence of imperfect substitution.

The coefficients of the interaction term in Table 1.4 indicate that arbitrage became more relevant in the post-2000 period. This observation aligns with the findings in the subsequent sections, which note that the index effect began to shrink after 2000. Furthermore, by focusing on the post-2000 period and comparing the coefficient of the interaction term with the baseline coefficient in column (1)[22], it is evident that a one standard deviation increase in arbitrage risk increases the price impact of demand shifts by approximately 30-45 percentage points of the baseline average effect. This explains the cross-sectional heterogeneity of stocks in the sensitivity of their prices to demand shifts.

Lastly, the coefficients of A_1 and A_2 are positive in all regressions, particularly highly significant in the post-2000 period. The interpretation of this coefficient is more nuanced. Note that if this coefficient had turned out to be zero, it would imply that the price change for two demand shifts that are otherwise equal except in direction would be symmetric as well, except again for the direction. However, in the presence of this statistically

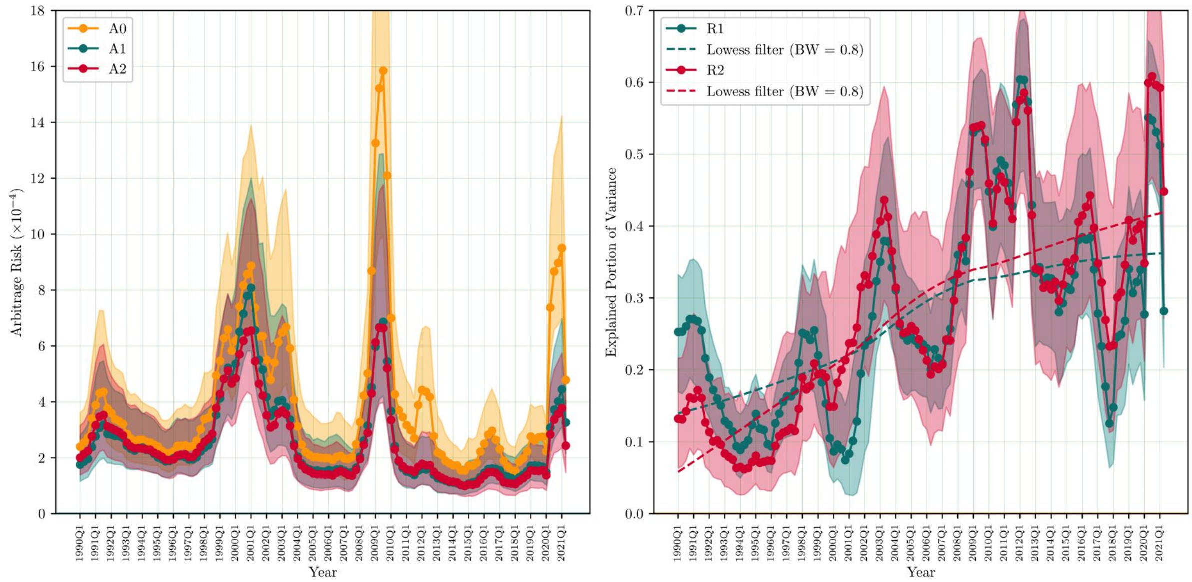
[22]If I re-estimate the baseline coefficient only for the post-2000 period, I still obtain an estimate of 0.487, statistically significant at the 1% level, which is very similar to the full sample estimate

positive coefficient on A_1 and A_2, these two otherwise equal opposite-direction demand shifts would have asymmetric price reactions such that the price reaction for the positive demand shift is larger in magnitude. This occurs because when the demand shift of index funds is positive, arbitrageurs must sell to the indexers. Yet, arbitrageurs do not naturally hold these stocks by default, so they have to short-sell to the indexers. Since short-selling is costly (Sikorskaya, 2023), they would be willing to take the other side of the index funds' trades only at a higher price. Therefore, the coefficients on A_1 and A_2 are associated with the asymmetric price reactions to opposite-direction demand shifts due to costly short-selling. Moreover, comparing the coefficients from the pre- and post-2000 periods indicates that this asymmetry due to short-selling costs has diminished over time. Appendix section A.6 studies the volatility reaction to index funds' demand shifts and there as well it is documented that increase in volatility is larger for the case of positive demand shifts. Appendix section A.5 uses four alternative definitions for the risk of arbitrage and verifies the finding of this section are robust to those alternative definitions.

Regression results in Table 1.4 show that one important determinant of the total price impact multiplier (when factoring MRF in the underlying regression equations (1.9) the coefficient is $\beta_0 + \beta_1 A^N_{k,jt}$) is the risk of arbitrage, which is also the only time-varying component in it. Figure 1.3 complements this by demonstrating how these measure has evolved over time. The left-hand side plot illustrates that arbitrage risk exhibited a declining trend post-2000 and stabilized at a low level, with the exceptions of the 2008 financial crisis and the COVID-19 pandemic in 2020.

The left-hand side plot in Figure 1.3 highlights two points: (i) arbitrage risk crucially depends on the stability of financial markets. This is not surprising given that arbitrage risk is a component of total risk that cannot be hedged. When total risk increases during episodes of financial distress, this unhedgeable component naturally rises to some extent as well. (ii) The overall reduction in arbitrage risk, alongside the results in Table 1.4,

Figure 1.3: Arbitrage Risk Over Time



This figure illustrates the evolution of stocks' arbitrage risk over time. Arbitrage risk measures are calculated as the variance of residuals from regressing one stock's excess return on the market's excess return over a window of [-365, -20] prior days (A_1), or on the excess return of the three closest industry-, size-, and book-to-market-matched "substitute" stocks (A_2). These substitutes are matched using quintiles of size and book-to-market, employing the 49 industry classifications from Fama and French (1997). For further details, refer to Wurgler and Zhuravskaya (2002). All lines represent the median, while the shaded areas of the same color indicate the 25th and 75th percentiles. On the left-hand side, the green line depicts the trend for A_1, and the red line shows the trend for A_2. The total variation of stock returns within the same window is represented in yellow. The right-hand side illustrates the portion of total stock price variation explained by substitutes, measured as total variance minus arbitrage risk divided by total variance. Lowess filters with a bandwidth of 0.8 are also displayed to smooth the graphs and highlight the trends. The sample period is (1990-01, 2021-06).

suggests that the decline in price impact in the latter part of the sample is partially attributable to decreased arbitrage risk. Although I do not have a long enough time series of the price impact multiplier to perform regression analysis, the current time series estimated in Figure 1.2 already exhibits a correlation coefficient of 0.71 with that of A_2. The parallel trends of arbitrage risk and the price impact multiplier in Figure 1.2, including the spikes in 2000, 2008 and 2020, reaffirm that the documented reduction in price impact is at least partially due to a decrease in arbitrage risk, rather than being an unrelated, continually decreasing time series.

The results in this section have demonstrated that the reduction in the overall price impact of stock prices in response to demand shifts is at least partially due to the overall reduction in stocks' arbitrage risk. However, a lingering question remains: what has led to this decrease in arbitrage risk? While a comprehensive answer to this question requires further research, I can address whether overall price fluctuations have decreased, or if better substitutes capture a larger portion of the variations.

The yellow line on the left-hand side of Figure 1.3 indicates that although overall variation has decreased in some episodes of the sample span, at the beginning and end of the sample period, it had similar magnitudes. However, the left-hand side plot shows a secular increasing trend in the portion of total variation explained by the substitutes. For instance, while the set of three substitutes used for A_2 were hedging 5-20 percent of the total variations in the 1990s, they can capture an average of about 42 percent of the total variations in the last decade of the sample. Thus, the reduction of arbitrage risk is indeed more due to better substitutability of stocks in cross-stock arbitrage than an overall reduction in volatility.

Implications for the Price Elasticity of Demand

The price elasticity of index funds' demand can be directly calculated as $-1/\beta$, in which β comes from the price impact estimates in the previous section. With flat or perfectly elastic demand curves, the price elasticity is $-\infty$, and with downward-sloping or imperfectly elastic demand, the price elasticity approaches zero.

Despite the crucial importance of this parameter in asset pricing models, the literature offers a very broad range of its estimates. In a recent study, Gabaix and Koijen (2022) lists the most recent estimates of price elasticity of demand. Using individual stocks characteristics like in the present paper, as opposed to those using factor level or macro estimates, the estimates range between -0.4 to -3.3. While some of these papers use index redefinition such as Pavlova and Sikorskaya (2022) and Chang et al. (2014), there are papers leveraging other demand shifts such as dividend payouts as in Schmickler (2020) and Hartzmark and Solomon (2022), and mutual fund flows in Lou (2012).

In assessing elasticities, my study centers on the price response throughout the entire period from announcement to implementation, denoted as CAR^{Eff}_{Ann} and presented in column (4) of Table 1.2. This approach offers the distinct advantage of incorporating the surprise effect of index decisions during the announcement while abstracting from market fluctuations. Such calculation reveals a price elasticity estimate of -2.5 within the full sample which increases largely in magnitude towards the recent era according to the time-varying model.

The estimates of elasticity in this section could be called into question if index funds, which incur transaction costs for their trades, do not rebalance their portfolio in response to changes in index composition. Similarly, the possibility that some active funds may trade based on index guidelines and so biasing the estimates upward can also be a concern. While these are legitimate identification concerns, they should not be overemphasized for two reasons. First, Pavlova and Sikorskaya (2022) have shown that passive funds benchmarked to S&P 500 have almost zero tracking error (0.2-0.4 % per annum) in their entire sample period and have held, on average, 99.6% of their benchmark stocks. This evidence suggests that at least passive funds benchmarked to S&P 500 do exactly as they are expected: They mirror the index portfolio with no discretion.

Second, although some active funds may trade based on index inclusion or exclusion decisions, it is unlikely that they would rebalance other stocks in that benchmark according to updated weights. Moreover, since they do not hold the entire index portfolio, their rebalancing would not occur in parallel with that of index funds. Furthermore, the bottom line of the paper, which will be examined in the subsequent sections, is that demand can explain the entire magnitude and the decreasing trend of the index effect in the post-2000 sample, leaving no room for information-based explanations. This takeaway will rely on the interaction of price impact estimates and the size of the demand shift. Thus, even if some active funds are closet indexers and hold a similar portfolio to index funds, meaning that my MRF measure is systematically understated because the true aggregate AUM is larger than what I have, this implies that the estimated price impact multipliers are systematically overstated due to the exact reciprocal of the same bias, and the interaction of the two remains the same quantity and the bottom line of the paper remains entirely valid.

Demand and Turnover

The weight surprises experienced by index incumbents due to the difference in the weight of added and deleted firms during reconstitutions are indeed relatively small in face value. However, this section shows they can induce a significant demand shift when multiplied by the large assets under the management of index funds[23].

To study the volume reaction of incumbents to demand shifts, I again focus on the entire period from the announcement to the effective day. As the outcome variable, I use first the cumulative turnover, where turnover is defined for each day as the ratio of the daily volume of trades to the number of shares outstanding, and second, the cumulative excess turnover, which is the same measure minus the average expected turnover of that stock for that number of days, with the average taken over a window of one month ending twenty days before the effective day. For the independent variable, I employ the absolute value of the MRF measure. This choice is motivated by the fact that index fund trades will increase the volume of trades regardless of the direction (buy or sell). To confirm this, I include a dummy variable indicating the sign of the MRF in the regressions which turns out zero.

Estimation results are reported in Table 1.5 using the regression equations (1.7) for full sample estimates and (1.8) to see temporal evolution. The coefficient estimates of $abs(MRF)$ are about 2.7 for total cumulative turnover, and the direction of demand

[23]As of 2022, the total assets under the management of passive index funds benchmarked against S&P 500 are estimated to be around USD 5.4 trillion. Index fact sheet, available on https://www.spglobal.com/spdji/en/indices/equity/sp-500/

shifts has negligible effect as predicted. This means that for every 1% of firms' shares demanded by index funds, 2.7% of firm shares will be traded in the relevant period from announcement to effective day. Obviously, not all these trades belong to index funds or are due to their demand shifts; there is a certain daily turnover for all stocks every day. To tease this out, I focus on cumulative excess turnover in columns (3) and (4), which takes the average daily turnover out of the dependent variable. These regressions show that for every 1% of shares outstanding demanded by index funds, there is a 0.92% of excess cumulative turnover from announcement to effective day. I also verify manually that this coefficient is not significantly different from 1 at the 95% confidence level. Intuitively, the entire demand of index funds manifests as excess turnover in this period, and only a slight portion of it is supplied by the usual liquidity available in the market. This also corroborates the findings in the previous section as it certifies much of the demand shifts are fed at the margin by the arbitrageurs.

Column (5) of Table 1.5 studies the heterogeneity of stocks in their reaction of excess cumulative turnover to demand shifts. Particularly, I sort stocks into quartiles based on their average turnover over a one-month window ending twenty days before the effective day and document that the increase in excess turnover in response to demand shifts is intuitively stronger for stocks that are unconditionally less liquid. Lastly, Columns (6) and (7) show that the reaction of turnover became larger over time.

Chinco and Sammon (2024) also study the excess turnover reaction of additions and deletions for multiple indices, concluding that the excess turnover on these stocks is double what can be explained by the assets under management of index funds. My coefficient estimate of 2.2 in the last decade for excess cumulative turnover in Column (7) complements their study by showing that, at least for the S&P 500, the excess turnover on index incumbents is also double what can be justified by passive index fund assets. In other words, the same closet indexers that do not openly identify as passive index funds yet hold index portfolio similar to passive index funds also inevitably have to rebalance index

Table 1.5: MRF and Turnover

	(1) CumTO	(2) CumTO	(3) $CumEx-TO$	(4) $CumEx-TO$	(5) $CumEx-TO$	(6) CumTO	(7) $CumEx-TO$
$abs(MRF)$	2.675***	2.681***	0.916***	0.918***			
	(11.37)	(11.38)	(10.17)	(10.21)			
$MRF>0$		0.000***		0.000**	0.000**		
		(3.20)		(2.24)	(2.54)		
$Q-AvgTO_M$=1 × $abs(MRF)$					1.306***		
					(21.73)		
$Q-AvgTO_M$=2 × $abs(MRF)$					0.853***		
					(7.30)		
$Q-AvgTO_M$=3 × $abs(MRF)$					0.969***		
					(6.39)		
$Q-AvgTO_M$=4 × $abs(MRF)$					0.450		
					(1.31)		
Period=Early × $abs(MRF)$						0.646***	0.003
						(9.42)	(0.05)
Period=Mid × $abs(MRF)$						3.367***	1.610***
						(3.81)	(6.23)
Period=Late × $abs(MRF)$						6.489***	2.194***
						(7.51)	(9.47)
N	289019	289019	289019	289019	289019	289019	289019
Adj. R-sq	0.372	0.372	0.131	0.131	0.131	0.373	0.132

This table reports the results of estimating panel regression in Equation (1.7) for index incumbents when the dependent variable is daily turnover. The sample includes all incumbent observations in the reconstitution days in the sample period (2000-01, 2021-06). The main independent variable is the absolute value of MRF_t^j, representing the absolute value of dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio. $MRF > 0$ is a dummy equal to one (zero) if index funds' demand shift is positive (negative), i.e. they should buy (sell). Quartiles of $AvgTO_M$ distinguish unconditional liquidity of different stocks. $AvgTO_M$ indicates the average turnover (volume of trades divided by the number of shares) of stocks over a one-month window ending 20 days priod to the effective dates. Period dummy variables indicate three sub-samples: Early (1990-1999), Mid (2000-2010), and Late (2011-2021).Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day) and the average cumulative excess turnover of all S&P 500 firms from announcement to effective day as a proxy of aggregate market liquidity around respective reconstitutions. All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: *p < 0.10; **p < 0.05; ***p < 0.01.

incumbents in the same way that index funds have to do. At the end, the sum of the portfolio weights must remain at 100%!

1.4.2 Index Additions and Deletions

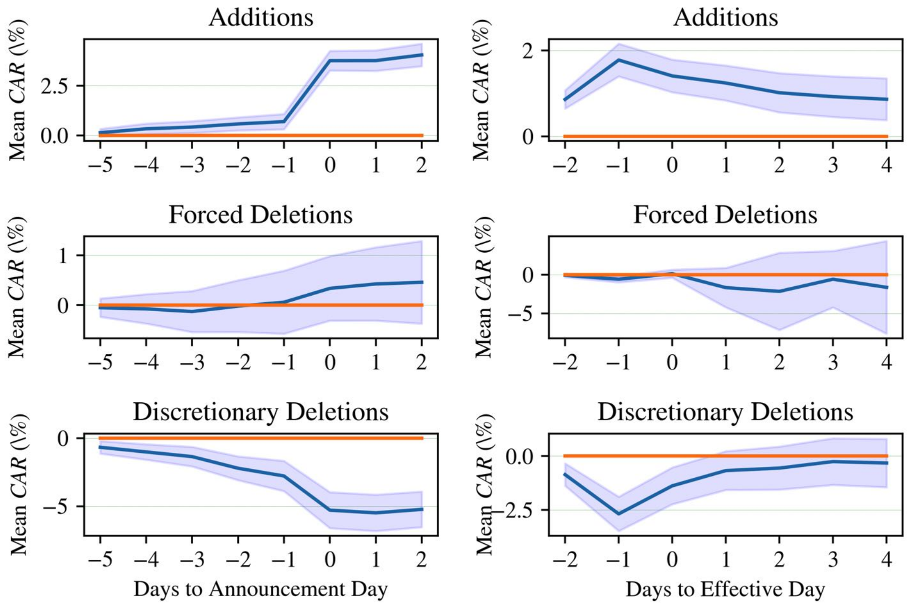
This section studies the price reactions to S&P 500 index reconstitutions by examining long-term and short-term price patterns of index addition and deletions. The first part of the section reports the abnormal return of added and deleted firms around the index reconstitution in an event study manner. The second part examines whether index funds' demand can justify the observed effect size by incorporating the estimates of price impact from Section 1.4.1.

Figures 1.4 and 1.5 show the abnormal returns experienced by the added and deleted firms around the index reconstitution, respectively, in the short and long time windows. The announcement day for each index reconstitution varies in the sample between 1 and 30 trading days before the effective day, with a high mass around the median of 5 trading days[24] which is one calendar week. Therefore, for the long window in Figure 1.5, I started calculating cumulative abnormal returns from 30 trading days before the effective day of events to ensure announcements happen within the window. Abnormal returns are also summarized numerically in Table 1.6. All abnormal returns are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations.

In the analysis, I distinguish between forced and discretionary deletions. I define forced deletions as deletions that were entirely delisted from the exchange within two week after their deletion to underscore that they are overshadowed by more significant events leading to delisting, rather than their removal from the index and the removal was not really a decision by index committee as these shares would cease trading shortly for another reason. Forced deletions exhibit stagnant abnormal returns until shortly before the effective day, yet these returns experience substantial swings in magnitude. Therefore, I refrain from commenting on their returns and instead focus on discretionary deletions,

[24] In vast majority of events, this distance is between 4 and 6 trading days.

Figure 1.4: Abnormal Returns of Index Additions and Deletions, short CARs



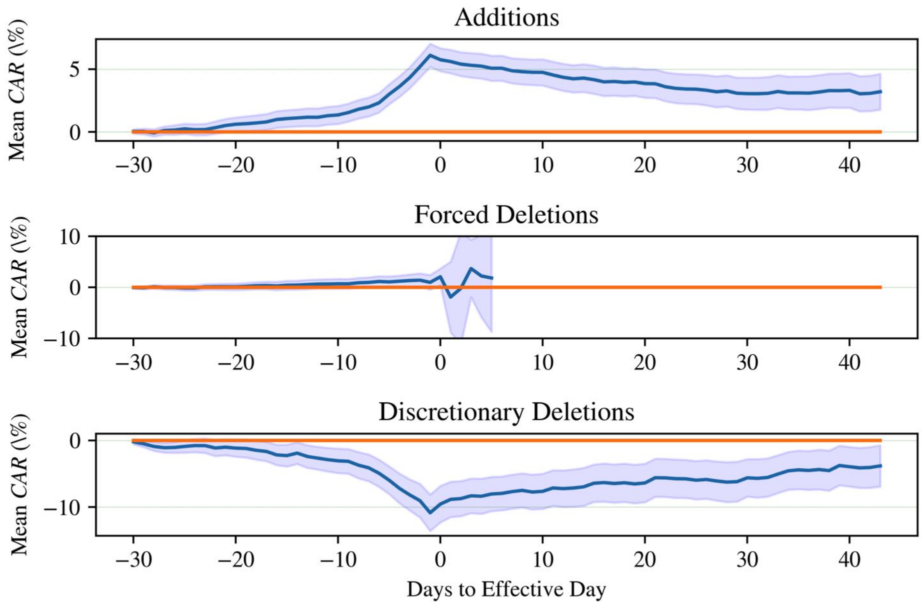
This figure presents the event study results for the cumulative abnormal returns of S&P 500 additions and deletions, distinguishing between forced and discretionary deletions. Forced deletions are those that were entirely delisted from the exchange within two week after their removal from the index. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. In the left figures day zero is the announcement day, and in the right figures it is the effective day. Vertical axes report outcomes in percentage. 95% confidence intervals are shown in light shadow around the means. The sample period is (1990-01, 2021-06).

which continue trading after being removed from the index.

For additions, the cumulative abnormal return in the period of $Ann-30$ to $Ann-2$ is positive and is significant. Between the three dates of announcement day and the days before and after it, only the announcement day has a significant return that averages to a whooping amounts of 3%. The period between announcement day and effective day, including both dates, has an average abnormal return of 3.9%. This is the main figure we know as *index effect*. Lastly, the cumulative abnormal returns reverts partially after the event.

Regarding discretionary deletions, the average CAR in the periods of $Ann-30$ to $Ann-10$ and $Ann-9$ to $Ann-2$ are both very large and significant summing to about

Figure 1.5: Abnormal Returns of Index Additions and Deletions, Long CARs



This figure presents the event study results for the cumulative abnormal returns of S&P 500 additions and deletions, distinguishing between forced and discretionary deletions. Forced deletions are those that were entirely delisted from the exchange within two week after their removal from the index. The event window begins 30 trading days before the effective day and ends 45 days after. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. Vertical axes report outcomes in percentage. 95% confidence intervals are shown in light shadow around the means. The sample period is (1990-01, 2021-06).

-6.2% although index announcement had not yet happened. This finding shows that the *index effect* is at best responsible for one-third of the trough we observe on Figure 1.5 on the effective day. The total period between announcement day and effective day, including both dates, exhibits an average abnormal return of -3.48. Similar to additions, the effect partially reverts afterwards.

Figure 1.6 provides a visual representation of index effect (CAR^{Eff}_{Ann}) for index additions and deletions in different years. This figure highlights an ostensibly puzzling trend: the magnitude of the index effect is decreasing over time, despite the significant increase in demand from index followers, as previously documented in Figure 1.1. This shrinking trend is also noted in Patel and Welch (2017), Bennett et al. (2023), and Greenwood and

Table 1.6: Abnormal returns of S&P 500 additions and deletions

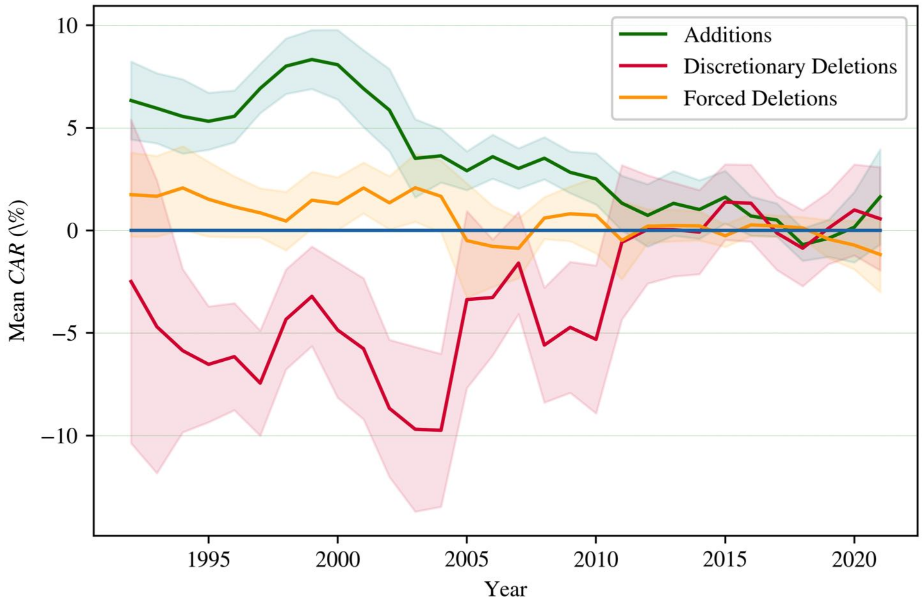
	Additions			Discretionary Deletions			Forced Deletions		
Time/Window	mean	pos (%)	N	mean	pos (%)	N	mean	pos (%)	N
CAR_{Ann-30}^{Ann-10}	1.27***	54.85	629	-2.89***	44.24	321	0.77	59.9	399
	3.18			-3.08			1.45		
CAR_{Ann-9}^{Ann-2}	0.96***	57.01	649	-3.28***	42.15	325	0.22	55.47	411
	3.87			-5.14			0.73		
AR_{Ann-1}	0.12	49.23	650	-0.57**	45.54	325	0.04	53.06	409
	1.26			-2.18			0.32		
AR_{Ann}	3.06***	79.23	650	-2.51***	29.85	325	0.1	52.53	396
	20.4			-8.59			0.78		
AR_{Ann+1}	0.0	48.31	650	-0.2	45.23	325	0.07	53.19	361
	0.04			-0.77			0.56		
CAR_{Ann}^{Eff}	3.88***	71.85	650	-3.48***	36.62	325	0.63***	56.06	396
	14.55			-6.66			2.63		
AR_{Eff-1}	0.92***	59.66	652	-1.82***	32.52	326	-0.54***	41.05	380
	5.93			-6.81			-3.34		
AR_{Eff}	-0.37***	43.56	652	1.3***	58.59	326	0.84***	60.37	328
	-3.96			4.66			4.8		
AR_{Eff+1}	-0.17*	48.16	652	0.71***	54.29	326	0.37	56.0	50
	-1.92			2.83			0.53		
CAR_{Eff+2}^{Eff+5}	-0.53***	46.17	652	0.63	50.31	324	-0.3	61.54	13
	-3.16			1.33			-0.16		
CAR_{Eff+6}^{Eff+10}	-0.32	45.55	652	0.42	50.46	325	-	-	-
	-1.61			0.8					
CAR_{Eff+11}^{Eff+20}	-0.82***	46.48	654	1.53**	52.02	321	-	-	-
	-3.3			2.5					

This figure presents the event study results for the cumulative abnormal returns of S&P 500 additions and deletions, distinguishing between forced and discretionary deletions. Forced deletions are those that were entirely delisted from the exchange within two week after their removal from the index. The first column indicates the related window or date. For windows of more than one day, the mean of cumulative abnormal returns over those intervals are reported, and for other rows that are single dates, the mean of respective abnormal return is shown. T-statistics for testing zero means are reported in parentheses. Columns indicated by "pos %" show the percentage of observations with positive values. N is the number of observations. Eff and Ann indicate effective and announcement days, respectively. Statistics are defined conditional on a minimum of ten observations. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. The sample period is (1990-01, 2021-06). Significance levels for a t-test of zero means are marked as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Sammon (2024). The decreasing magnitude of the index effect is reflected in professional publications even a decade earlier than in academic ones (see, for instance, Soe and Dash (2008) and the web articles that referred to it).

Two stylized facts observed collectively in Table 1.6, Figure 1.4, and Figure 1.6 are

Figure 1.6: Cumulative Abnormal Returns of Additions and Deletions Over Time



The figure shows the three-year moving average of cumulative abnormal return of S&P 500 additions (green line), forced deletions (orange line), and discretionary deletions (red line) by year. Forced deletions are those that were entirely delisted from the exchange within two week after their removal from the index. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. Vertical axes report outcomes in percentage. 95% confidence intervals are shown in light shadow around the means. The sample period is (1990-01, 2021-06).

worth mentioning. First, long-term abnormal returns for additions and discretionary deletions both seem long-lasting. Second, the CAR^{Eff}_{Ann} in Table 1.6 is smaller in magnitude than the peak CAR in Figure 1.4 for both additions and discretionary deletions. It means that a significant portion of abnormal returns in peak CARs have been realized before the index announcements which could have influenced the index committee choice of these firms.

The preceding section demonstrated that the price impact multiplier has exhibited a declining trend over the past decade, independent of and beyond index decisions. This phenomenon can potentially explain the shrinking magnitude of the index effect despite the increasing size demand shocks, especially if it has evolved more rapidly than the growth of passive investing. In light of these empirical observations, the remainder of this section formally establishes that a demand model can effectively replicate both the magnitude of the index effect and its decreasing trend.

To provide a benchmark for appropriate price reaction to demand shifts in the absence of information, I use the price impact regressions in Section 1.4.1 to find the counterfactual price reaction of index additions and deletions if those events were truly mechanical (not influenced by informational component) and examine if such predictions are in the same magnitude of actual figures. For the model with a time-invariant coefficient estimate, I produce a predicted return ($C\hat{A}R_i$) for each added or deleted stock based on its MRF and the estimates of coefficients in regression (4) of Table 1.2 whose outcome variable is CAR^{Eff}_{Ann}. Regression (4) is chosen since it measures the price reaction precisely on the same time interval for index incumbent that we measure CAR^{Eff}_{Ann} on for additions and deletions. I produce the predicted value based on the model with a time-varying coefficient in a similar manner. Crucially, in both models, I incorporate the same controls and fix effects in making the predicted values that were used in the reference estimation. Results of this prediction exercise are presented and discussed in multiple manners in what follows. Appendix section A.9 verifies the linearity of price impact estimates in response to demand shifts and discusses its implications.

Table 1.7 displays actual cumulative abnormal returns (CAR^{Eff}_{Ann}) for additions and discretionary deletions, as well as the difference between actual and counterfactual values ($C\hat{A}R^{Eff}_{Ann}$) using both the time-invariant model and a time-variant model for price impact. These differences are, in fact, the residuals of the underlying predictive regression using information-free demand shifts and can be attributed to the informational components in the index effect that were discussed in the literature review. If these residuals turn out to be significantly different from zero, this would provide evidence for a material informational component in the index effect. The table first presents full sample results, followed by results for the post-2000 period.

In the full sample, both models fail to capture the mean cumulative abnormal re-

Table 1.7: Comparison of Actual and Predicted Abnormal Returns

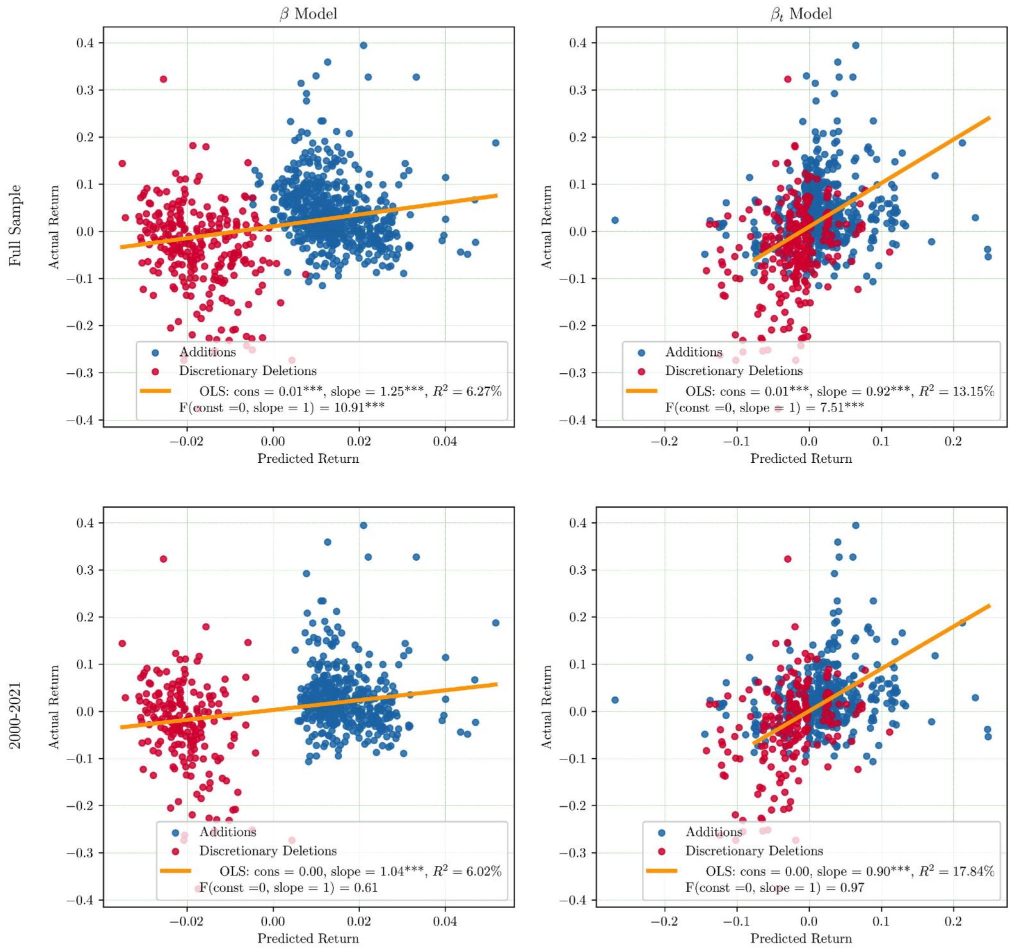
Period	Additions/Deletions	CAR	$CAR - \hat{CAR}$	
			β model	β_t model
full sample	Additions	3.92***	2.52***	2.18***
		(0.27)	(0.28)	(0.28)
	Discretionary Deletions	-3.62***	-1.89***	-1.80***
		(0.55)	(0.56)	(0.52)
post 2000	Additions	2.64***	0.95**	0.46
		(0.30)	(0.31)	(0.30)
	Discretionary Deletions	-3.13***	-1.09	-0.97
		(0.66)	(0.67)	(0.61)

This table compares the sample means of the counterfactual and actual index effects for additions and discretionary deletions. Discretionary deletions are index deletions that are not delisted from their exchange at least within the next two weeks after their removal from the index. The actual index effect is defined as the cumulative abnormal return of these stocks from the announcement to the effective day of the changes. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. The counterfactual index effect represents the predicted value for the actual index effect based on MRF and estimates of price impact (or elasticities) derived from the sample of incumbent stocks, as discussed in Section 1.4.2. Please refer to the main text for detailed procedural information. All returns are expressed as percentages. The model used for these predictions is reported in the top row. Each model is estimated on the sample of index incumbents and then employed to generate the predicted values for additions and deletions using same fix effects and controls. The sample period is (1990-01, 2021-06). Significance levels for a t-test of zero means are marked as: $^{*}p < 0.10; ^{**}p < 0.05; ^{***}p < 0.01$.

turns, with residuals significantly positive for additions and significantly negative for discretionary deletions. This indicates that the test cannot reject the existence of information over the full sample. In the post-2000 period, the model with a fixed price impact multiplier still cannot reject the existence of information, as the residuals remain large-approximately one third of the entire index effect size-and significant for additions. However, when the model with a time-varying multiplier is employed, the residuals are smaller and statistically insignificant. This outcome suggests that demand-based explanations alone are sufficient to account for the magnitude of abnormal returns observed with S&P 500 additions and deletions. However, if the researcher fails to consider the time-varying price impact (or elasticities), she would not be able to explain one third of the effect size, and significant residuals might erroneously be attributed to information.

It is important to emphasize that the findings mentioned above do not entirely rule out the relevance of all information-based explanations, such as information dissemination,

Figure 1.7: Actual VS. counterfactual index effect



The figure shows the counterfactual (x-axis) and actual index effects (y-axis) for additions in blue and discretionary deletions in red dots. Discretionary deletions are index deletions that are not delisted from their exchange at least within the next two weeks after their removal from the index. The actual index effect is defined as the cumulative abnormal return of these stocks from the announcement to the effective day of the changes. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. The counterfactual index effect represents the predicted value for the actual index effect based on MRF and estimates of price impact (or elasticities) derived from the sample of incumbent stocks, as discussed in Section 1.4.2. Please refer to the main text for detailed procedural information. In the left figure predicted values are made using a static (time-invariant) model, while in the right figure a dynamic model with time varying price impact multiplier is used. The orange line shows the OLS regression of actual values on the counterfactual ones. F-statistics for a joint hypothesis of a zero intercept and a slope of one is reported in the legends. The sample period is (1990-01, 2021-06).

investor awareness, liquidity, and attention. Indeed, these factors may individually have some heterogeneous impact on the affected stocks. Instead, the results suggest that if these information-related components do exist, their effects are either rather long-term or offset each other in the short run, resulting in a relatively minimal net marginal impact on the average magnitude of abnormal returns post-2000. This is a plausible scenario, given the mixed evidence regarding the influence of these components on stocks. For instance, while improved awareness among market participants may have a positive effect on additions, as noted in Chen et al. (2004), Bennett et al. (2023) finds that increased scrutiny following stock additions to the S&P 500 can have adverse effects on a firm's performance. My methodology has the unique advantage over existing works in that it leaves room for an aggregate informational component without taking a side on its true nature, and then shows that although a combination of those components was materially at work to explain the index effect as documented in earlier works, they no longer function as such in the most recent years.

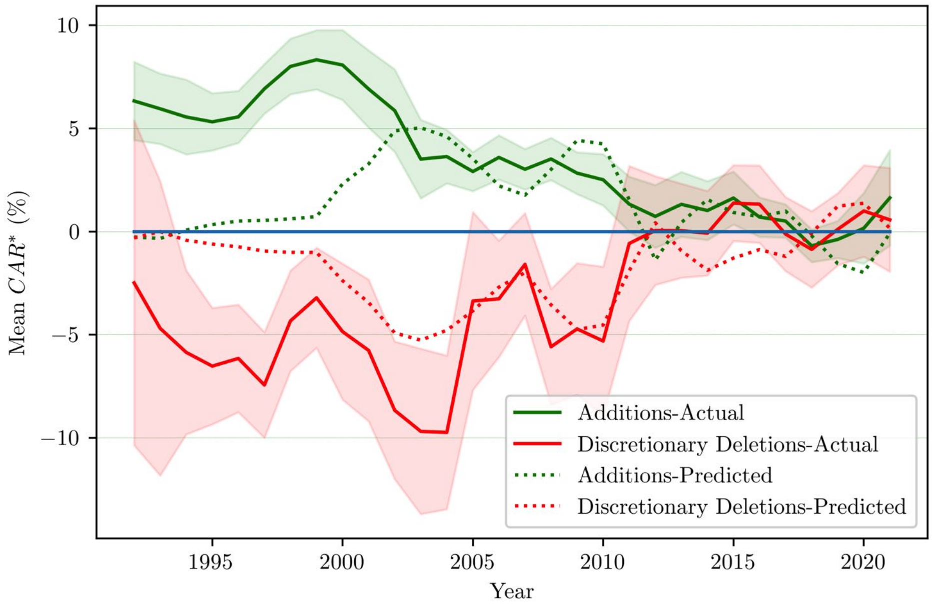
Sections A.7 and A.8 in the Appendix provide additional evidence against the hypothesis that S&P 500 index decisions convey information, as they examine the long-term returns and fundamentals associated with index additions and deletions in the post-2000 period. These sections yield two noteworthy observations: Firstly, neither calendar-time portfolios comprising recent index additions nor those composed of recent index deletions yield sustained *alpha*s. Secondly, additions to the index typically occur when firms are at their best fundamental stance rather than when they have the highest growth potential. Conversely, deletions occur when firms report a negative net income following a trend of declining performance. These results suggest mean reversion in the firms' performance, and show that index decisions seem rather retrospective than forward-looking in the post-2000 period.

Figure 1.7 displays a scatter plot of the actual and information-free counterfactual index effect based on the two models. Each point on the graph represents a stock addition or discretionary deletion, with its actual CAR^{Eff}_{Ann} on the vertical axis and its counterfactual $C\hat{A}R^{Eff}_{Ann}$ value on the horizontal axis. The figure also details the simple line from a univariate OLS regression. The aim of these scatter plots is to test a joint hypothesis: there is no information in index effect (i.e. constant of OLS is zero, such that the mean of the actual values is entirely captured by the counterfactuals) and correct model specification (i.e. slope is equal to one, such that for each unit of increase in the actual index effect there is an equivalent one unit of increase in the counterfactuals). The F-statistic for this joint test is reported in each graph. In the full sample, the F-test rejects the hypothesis, as even visually the constant is not zero and the slope is not so close to one. In the post-2000 period, however, the test is not rejected in either of the models, which provides further assurance for no information in index effect in this era and correct model specification.

Figure 1.7 also reports the R^2 of the underlying regressions, which by definition are *cross-sectionally* out of sample since the added and deleted stocks were not included in the sample of the training models that were estimated on the index incumbents. The R^2 of the model with time-varying coefficients is significantly higher, reaching an impressive amount of 18% for the post-2000 period. This is because this model more effectively captures the variation in price reactions due to changes in market conditions. The predictions of the time-varying model have a wider range in which both added and deleted stocks exhibit negative and positive returns, just as in the actual returns.

So far, the findings of this section show that in the post-2000 era, index funds' demand shifts can explain the entire magnitude of the index effect. Figure 1.8 completes the picture by showing that the decreasing trend of the index effect is also explained by these demand shifts. The figure presents a three-year moving average of the index effect for additions and discretionary deletions over time, along with the three-year moving average of its counterfactual counterpart. We observe, once again, that in the 1990s, the small amount of money invested in index funds cannot realistically explain the huge abnormal returns that firms were experiencing upon additions or removals from the index, thereby certifying another (information-related) component in the index effect. However, starting

Figure 1.8: Actual VS. counterfactual index effect over time



The figure shows the evolution of counterfactual (dashed lines) and actual index effects (solid lines) for additions in green and discretionary deletions in red over time. Discretionary deletions are index deletions that are not delisted from their exchange at least within the next two weeks after their removal from the index. The actual index effect is defined as the cumulative abnormal return of these stocks from the announcement to the effective day of the changes. Abnormal returns are calculated using Carhart (1997) four factor model, whose parameters are estimated over a window of one-year ending 20 days prior to the observation day conditional on a minimum of 100 observations. The counterfactual index effect represents the predicted value for the actual index effect based on MRF and time-varying estimates of price impact on index incumbents discussed in Section 1.4.2, and shown in Figure1.2. Please refer to the main text for detailed procedural information. The sample period is (1990-01, 2021-06).

from around 2000, the counterfactuals fall within the 95% confidence interval of the actual values, and they also exhibit a decreasing trend similar to that of the actual ones.

Cai (2007) provides one of the most important evidence in favor of the information hypothesis on S&P 500 reconstitutions by examining the price and volume reactions of industry- and size-matched firms. He focuses exclusively on index additions and uses a sample of additions from 1976 to 2001, finding a significantly positive price reaction but no volume reaction for the matching firms. This suggests that the index addition conveys favorable information not only about the added firm and but also about other firms in its industry. In the last part of this section I partially replicate his results which focuses

Table 1.8: Index effect and industry-matched incumbents

	(1) CAR_{Ann}^{Eff}	(2) CAR_{Ann}^{Eff}	(3) CAR_{Ann}^{Eff}	(4) CAR_{Ann}^{Eff}	(5) CAR_{Ann}^{Eff}
SameSIC3Add	0.001**				
	(2.49)				
Post-2000=0 × SameSIC3Add		0.003**			
		(2.34)			
Post-2000=1 × SameSIC3Add		0.001			
		(1.34)			
SameSIC3Del			-0.001		
			(-1.61)		
Post-2000=0 × SameSIC3Del				0.003	
				(1.44)	
Post-2000=1 × SameSIC3Del				-0.003***	
				(-2.68)	
Post-2000=0 × SameSIC3Del × Discretionary					-0.007**
					(-1.98)
Post-2000=1 × SameSIC3Del × Discretionary					0.001
					(0.73)
Controls	Y	Y	Y	Y	Y
FE	Y & S	Y & S	Y & S	Y & S	Y & S
N	288997	288997	288997	288997	288997
Adj. R-sq	0.006	0.006	0.006	0.006	0.006

This table reports the results of estimating panel regression in Equation (1.7) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include dummy variables if the incumbents are in the same three digit SIC industry with additions or deletions. Discretionary is a dummy variable that is equal to one if there is a discretionary deletion in respective reconstitution and zero otherwise. Post-2000 is dummy variable indicating year 2000 or afterwards. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

directly on information.

To study the existence of the information channel following the methodology in Cai (2007), I regress CAR_{Ann}^{Eff} for incumbents on dummy variables that equal one if an incumbent is in the same industry as the additions and deletions. Following Cai (2007), I define industry using two-digit or three-digit SIC codes. Table 1.8 shows the estimation results using dummies for the same three-digit SIC industry, and Table A.13 in Appendix section A.10 replicates it for two-digit SIC codes, drawing a similar picture. Column (1) in Table 1.8 shows that in the full sample, incumbents in the same industry as additions experience 10 bps more cumulative abnormal return compared to other incumbents. Column (2) uses a dummy to separate pre-2000 and post-2000 and shows that this effect only belongs to the pre-2000 period, where same-industry firms experience 30 bps more cumulative abnormal return, while in the post-2000 period, this effect does not continue. The estimate in the pre-2000 era is also qualitatively similar to that of Cai (2007). So far, I have shown that: (i) I can replicate the evidence of information in index additions documented earlier by Cai (2007) for his sample period[25], and (ii) this effect no longer exists in the post-2000 period.

Cai (2007) does not study the information effect on S&P 500 deletions. One reason for this could be the absence of distinguishing between forced and discretionary deletions; if the information exists, it could go either way for forced deletions. For example, if a forced deletion occurs because a firm is being delisted as a target of a merger and acquisition, it could represent potentially favorable information. Conversely, if it is being liquidated in bankruptcy, it is definitely bad news. Distinguishing between forced and discretionary deletions enables me to extend the results in Cai (2007) and focus only on discretionary deletions because, for forced deletions, much more significant events are happening than simply being removed from the index.

Column (3) of Table 1.8 shows that in the full sample there doesn't seem to be any information effect for incumbents in the same industry as deletions. However, once I focus on the pre- and post-2000 periods separately in Column (4), there is a negative effect on

[25]I could also use 1990-2001 and post-2001 to match the sample period to the end of the one in Cai (2007), and nothing would materially change, so I prefer to stick to the pre-2000, post-2000 division to keep it consistent with the rest of the paper

incumbents in the same industry as deletions in the post-2000 period. When I narrow down to the deletion effect by interacting a dummy that distinguishes discretionary deletions in Column (5), a consistent picture appears again: discretionary deletions conveyed negative information for firms in the same industry in the pre-2000 period, which was also very large on average-70 bps-and this effect disappears in the post-2000 period. Appendix section A.10 verifies that the demand effect documented earlier is entirely independent of the information effects documented here. In other words, the magnitude of the price impact multiplier is unchanged once I control for the same-industry dummies discussed here.

1.5 Conclusion

While it is widely recognized that changes in index composition can lead to substantial abnormal returns for affected stocks, the exact origin of these outcomes remains unclear. Evidence linking information effects (Denis et al., 2003; Cai, 2007) and empirical studies showing changes in firm behavior and market attention following index additions (Chen et al., 2004; Bennett et al., 2023) raise questions about whether index funds' demand shifts alone can fully account for these abnormal returns or if information-based factors unrelated to these demand shifts also play a significant role in index effect for indices that are not rule-based, such as S&P 500.

In this paper, I leverage variations in stocks' portfolio weights resulting from non-fundamental sources—specifically, differences in the size and portfolio weight of index additions and deletions. By quantifying these mechanical demand shifts, this study demonstrates that passive index funds' rebalancing exerts a substantial impact on stock prices, liquidity, and volatility. Focusing on index incumbents rather than actively involved stocks, I measure the aggregate and temporal price impact of index funds' demand independently of informational components. The results indicate that a 1% shift in index funds' demand, unrelated to information or fundamentals, results in a significant price change of approximately 40 basis points between the announcement and implementation of stock composition changes. This effect was most strong around the turn of the century, and decreased afterwards, which is termed here as the flattening of demand curves.

Furthermore, I demonstrate that arbitrage plays a crucial role in efficiently pricing stocks. Stocks with closer substitutes experience smaller price movements in response to demand shifts, as arbitrageurs can more effectively feed to these shifts with a lower risk stemming from offloading part of the risk on substitutes. Finally, I show that the reduction in price impact observed earlier comoves with the overall arbitrage risk due to closer substitutability, contributing to market efficiency.

Building on these insights, I have investigated whether demand alone can account for the magnitude of abnormal returns experienced by S&P 500 additions and deletions. I demonstrate that: (i) in the pre-2000 era, the index effect was too large to be reasonably attributed solely to the demand shifts of index funds, given their small size relative to firms' market capitalization; therefore, an information-based component was indeed present in the S&P 500 index effect, and (ii) in the post-2000 era, both the magnitude of the index effect and its decreasing trend can be explained using only a demand-based model that features time-varying price impact multiplier, with no additional information-based explanations required.

Appendix A

Appendices

A.1 An example of incumbent weight changes

This section offers a simple example to illustrate the dynamics of index incumbent weight changes when there is a heterogeneity in the weight of added and deleted firm(s). Assume some firm X with a free-float market value of $FFMV_X$ is going to replace firm Y with a free-float market value of $FFMV_Y$ in the composition of S&P 500. Suppose day $t+1$ is the effective day this index reconstitution and the free-float market value of all index incumbents at the close of day t equals $FFMV_{incs}$. Let firm A be a representative index incumbent. Therefore A has already been in the index before the reconstitution and also remains in it afterwards. Table A.1 compares the weight and size outcomes before and after the reconstitution.

Suppose that the reconstitution results in a 1% increase in the total $FFMV$ of index, that is $(FFMV_{incs} + FFMV_X) = (1.01) * (FFMV_{incs} + FFMV_Y)$. The change in the weight of A will be

Table A.1: Changes happening in a reconstitution event

	At the close of day t	Before the open of day $t+1$
Constituents	All incumbents + Y	All incumbents + X
Total $FFMV$ of index	$FFMV_{incs} + FFMV_Y$	$FFMV_{incs} + FFMV_X$
w^X	0	$\frac{FFMV_X}{FFMV_{incs}+FFMV_X}$
w^Y	$\frac{FFMV_Y}{FFMV_{incs}+FFMV_Y}$	0
w^A	$\frac{FFMV_A}{FFMV_{incs}+FFMV_Y}$	$\frac{FFMV_A}{FFMV_{incs}+FFMV_X}$

This tables summarises the changes that happen in a reconstitution day. X is a firm that is added to the index in this day, Y is the firm that is dropped from it, and A is an already index constituent than continues to remain in the index list. Dat $t+1$ is the effective day of this reconstitution.

$$\Delta w^A_{t+1} = w^A_{\text{before open } t+1} - w^A_{\text{at the close } t} = \frac{FFMV_A}{FFMV_{incs} + FFMV_X} - \frac{FFMV_A}{FFMV_{incs} + FFMV_Y}$$
$$= w^A_{\text{at the close } t}(\frac{1}{1.01} - 1) \simeq (-0.01) w^A_{\text{at the close } t}$$

Therefore, in a nutshell, when a firm is added to the S&P 500 and achieves 1% more weight in the index than the firm it has replaced, all other firms in the index will lose about 1% of their weight in the index after this reconstitution.

Note that since this non-market-driven action has changed the index $FFMV$ overnight, the index divisor will be adjusted to avoid any jump in the index level. In this simple example, index $FFMV$ has increased 1% overnight. So the index divisor will be increased exactly 1% as well. Therefore the last ratio in the equation (1.4) of Section 1.2 will be:

$$\frac{Divisor_t}{Divisor_{t+1}} = 0.99$$

and this is precisely the channel through which the difference in the weight of additions and deletions spills over the weight of index incumbents.

A.2 Alternative Regression Specifications

Table A.2 replicates the results from Table 1.2 for the period after 2000, during which I can utilize day fixed effects instead of year fixed effects. In the pre-2000 period, I cannot employ day fixed effects because weight surprises on the incumbents are estimated solely by the last term in Equation (1.4), which represents the inverse of divisor growth—a common factor for all incumbents each day that would be entirely absorbed by the day fixed effects. Table A.3 replicates the results without firm controls. Table A.4 performs the same analysis but uses an alternative division of the sample period into early, mid, and late sub-samples. All these alternative specifications show the robustness of results in Table 1.2.

A.3 Discussion of Theory with Multiple Substitutes

This section extends the findings of Gromb and Vayanos (2010) to scenarios involving multiple substitute assets, drawing a direct comparison to the methodology employed in this study. Gromb and Vayanos (2010) utilize a two stock model to quantify the pricing implications of imperfect substitution, which curtails arbitrageurs' ability to absorb information-free demand shocks without affecting prices. In their model, Stock A experiences a non-zero demand shock u, potentially shifting its price away from the expected

Table A.2: MRF and Price Impact-Including Day Fixed effects

Dependent Variable	(1) Ret_{Eff}	(2) Ret_{Eff-1}	(3) Ret^{open}_{Eff}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff-2}_{Ann}	(6) CAR^{Eff}_{Eff-1}	(7) AR_{Eff}	(8) AR_{Eff-1}
				Mid Sample (2000-2010)				
MRF	0.299	0.524**	0.145	1.006***	0.348**	0.658**	0.339*	0.318**
	(1.60)	(2.12)	(1.15)	(3.08)	(2.27)	(2.32)	(1.81)	(2.48)
Adj. R-sq	0.401	0.386	0.342	0.018	0.018	0.016	0.016	0.013
N	118685	118685	118676	118676	118676	118676	118685	118676
				Late Sample (2011-2021)				
MRF	0.115	-0.151	0.100*	-0.077	-0.035	-0.042	0.154	-0.196*
	(0.97)	(-1.23)	(1.80)	(-0.40)	(-0.24)	(-0.22)	(1.35)	(-1.68)
Adj. R-sq	0.421	0.347	0.583	0.019	0.021	0.019	0.020	0.018
N	90290	90290	90290	90286	90286	90286	90290	90286
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Stock-FE	Y	Y	Y	Y	Y	Y	Y	Y
Day-FE	Y	Y	Y	Y	Y	Y	Y	Y

This table presents the results of estimating the panel regression specified in Equation (1.7) for index incumbents. Results are produced separately for two sub-samples: Mid (2000-2010), and Late (2011-2021). The Early sample is not included because for that period the day fixed effects absorb the effect of MRF as due to data availability only the divisor growth is used in making the MRF measures of demand shifts. Please discuss the main text for more details on data preparation. The independent variable is MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and day fixed effects. All abnormal returns are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations. T-statistics based on standard errors clustered by stocks are reported in parentheses. Significance levels are denoted as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.3: MRF and Price Impact-Excluding Firm Controls

Dependent Variable	(1) Ret_{Eff}	(2) Ret_{Eff-1}	(3) Ret^{open}_{Eff}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff-2}_{Ann}	(6) CAR^{Eff}_{Eff-1}	(7) AR_{Eff}	(8) AR_{Eff-1}
				Full Sample (2000-2021)				
MRF	0.477***	0.693***	0.234***	0.383***	0.163**	0.220***	0.140**	0.080*
	(7.14)	(8.70)	(5.88)	(3.58)	(2.13)	(2.73)	(2.51)	(1.73)
Adj. R-sq	0.016	0.021	0.016	0.004	0.003	0.002	0.001	0.002
N	308750	308750	295250	308726	308726	308726	308750	308726
				Early Sample (1990-1999)				
MRF	0.759***	0.970***	0.248***	0.345***	0.184*	0.162**	0.063	0.100*
	(13.57)	(16.25)	(9.87)	(2.79)	(1.87)	(2.24)	(1.13)	(1.81)
Adj. R-sq	0.012	0.013	0.007	0.011	0.009	0.004	0.003	0.002
N	93755	93755	80681	93746	93746	93746	93755	93746
				Mid Sample (2000-2010)				
MRF	-0.143	0.693***	0.112	0.950***	0.359**	0.591**	0.295	0.296**
	(-0.89)	(2.66)	(0.92)	(3.06)	(2.29)	(2.22)	(1.61)	(2.51)
Adj. R-sq	0.015	0.022	0.016	0.006	0.005	0.002	0.001	0.002
N	124702	124702	124276	124692	124692	124692	124702	124692
				Late Sample (2011-2021)				
MRF	0.318*	-0.127	0.334***	-0.075	-0.077	0.002	0.203*	-0.201*
	(1.71)	(-1.10)	(3.41)	(-0.39)	(-0.53)	(0.01)	(1.65)	(-1.68)
Adj. R-sq	0.020	0.005	0.021	0.007	0.004	0.004	0.002	0.006
N	90290	90290	90290	90286	90286	90286	90290	90286
Controls	N	N	N	N	N	N	N	N
Stock-FE	Y	Y	Y	Y	Y	Y	Y	Y
Year-FE	Y	Y	Y	Y	Y	Y	Y	Y

This table presents the results of estimating the panel regression specified in Equation (1.7) for index incumbents without firms controls. The full sample comprises all incumbent observations on reconstitution days throughout the entire sample period (1990-01 to 2021-06). Results are then replicated separately for three sub-samples: Early (1990-1999), Mid (2000-2010), and Late (2011-2021). The bottom rows assess the statistical significance of the differences between sub-sample estimates. The independent variable is MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio. All regressions include stock fixed effects, and year fixed effects. All abnormal returns are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations. T-statistics based on standard errors clustered by stocks are reported in parentheses. Significance levels are denoted as: *p < 0.10; **p < 0.05; ***p < 0.01.

Table A.4: MRF and Price Impact-Alternative Division of Sample Period

Dependent Variable	(1) Ret_{Eff}	(2) Ret_{Eff-1}	(3) Ret^{open}_{Eff}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff-2}_{Ann}	(6) CAR^{Eff}_{Eff-1}	(7) AR_{Eff}	(8) AR_{Eff-1}
				Early Sample (1990-1999)				
MRF	0.869***	1.150***	0.254***	0.348**	0.175	0.174**	0.059	0.116**
	(14.23)	(18.75)	(10.05)	(2.50)	(1.57)	(2.22)	(0.98)	(2.05)
Adj. R-sq	0.010	0.017	0.007	0.017	0.017	0.004	0.005	0.003
N	80042	80042	79189	80033	80033	80033	80042	80033
				Mid Sample (2000-2009)				
MRF	-0.085	0.797***	0.132	1.103***	0.457***	0.647**	0.303	0.344***
	(-0.55)	(2.63)	(1.05)	(3.27)	(2.87)	(2.28)	(1.62)	(2.62)
Adj. R-sq	0.017	0.032	0.016	0.010	0.008	0.004	0.004	0.005
N	112835	112835	112826	112826	112826	112826	112835	112826
				Late Sample (2010-2021)				
MRF	0.202	-0.139	0.280***	-0.106	-0.076	-0.031	0.178	-0.209*
	(1.15)	(-1.23)	(3.05)	(-0.55)	(-0.54)	(-0.16)	(1.53)	(-1.79)
Adj. R-sq	0.019	0.010	0.025	0.009	0.007	0.004	0.004	0.007
N	96135	96135	96135	96130	96130	96130	96135	96130
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Stock-FE	Y	Y	Y	Y	Y	Y	Y	Y
Year-FE	Y	Y	Y	Y	Y	Y	Y	Y

This table presents the results of estimating the panel regression specified in Equation (1.7) for index incumbents for an alternative division of sample periods into three sub-periods. The full sample comprises all incumbent observations on reconstitution days throughout the entire sample period (1990-01 to 2021-06). Results are then replicated separately for three sub-samples: Early (1990-1999), Mid (2000-2010), and Late (2011-2021). The bottom rows assess the statistical significance of the differences between sub-sample estimates. The independent variable is MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. All abnormal returns are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations. T-statistics based on standard errors clustered by stocks are reported in parentheses. Significance levels are denoted as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

payoff $\bar{d_A}$. Arbitrageurs in their framework utilize Stock B as the sole substitute asset to correct mispricing, albeit imperfectly when assets are not perfect substitutes.

They establish that in equilibrium, the price of Asset A can be expressed as:

$$p_A = \bar{d_A} + \alpha\sigma_A^2(1-\rho^2)u \tag{A.1}$$

where initially the steady-state price equals the intrinsic value $\bar{d_A}$. This formulation implies that the price impact of a demand shock u is

$$\frac{p_A - \bar{d_A}}{\bar{d_A}} = \alpha\sigma_A^2(1-\rho^2)\frac{u}{\bar{d_A}} \tag{A.2}$$

This relationship highlights several key insights: Firstly, larger demand shocks exert a greater price impact ($\partial p_A/\partial u > 0$) as arbitrageurs require higher compensation for bearing increased risk. Secondly, a given demand shock u induces a more pronounced price impact when arbitrageurs exhibit greater risk aversion (α) and when Asset A's payoff uncertainty (σ_A) is heightened, as both conditions necessitate higher risk premiums. Moreover, the shock's impact intensifies when Asset A's payoff is less correlated with Asset B's ($|\rho|$), reducing arbitrageurs' hedging effectiveness with Asset B. Thus, assets characterized by higher idiosyncratic risk and fewer substitutes exhibit heightened sensitivity to demand shocks. I want to show that the part $\sigma_A^2(1-\rho^2)$ is equivalent to the arbitrage risk measure used in this study in the case of multiple assets.

Now, consider running an OLS regression of Asset A's return on that of Asset B, as follows:

$$r_t^A = \beta r_t^B + \varepsilon_t \tag{A.3}$$

In this setting, the expected values of the regression coefficient and R^2 are given by:

$$E[\beta] = \frac{cov(r^A, r^B)}{var(r^B)} = \rho\frac{\sigma_A}{\sigma_B} \tag{A.4}$$

$$E[R^2] = \beta^2 var(r_t^B)/var(r_t^A) = \rho^2 \frac{\sigma_A^2}{\sigma_B^2} * \frac{\sigma_B^2}{\sigma_A^2} = \rho^2. \tag{A.5}$$

Therefore, the variance of the residuals in such a regression is exactly: $var(\varepsilon) = (1 - R^2) * var(r^A) = (1 - \rho^2)\sigma_A^2$. This quantity represents the total variance of a long position in Asset A (σ_A^2) minus what the arbitrageur can hedge using an asset with only a correlation ρ, that is $\rho^2\sigma_A^2$.

This simple derivation shows that the coefficient in the price impact multiplier in the two-stock model of Gromb and Vayanos (2010) is equivalent to the variance of residuals in a regression of the impacted asset's return on its substitute's return, which in their model is only asset B. However, the intuition can be extended to cases with multiple substitutes. In such instances, arbitrageurs still bear residual risk even after forming the optimal linear combination of substitutes as their hedging position. This residual risk can similarly be quantified by the variance of residuals from an OLS regression of the impacted asset's return on the returns of its substitutes, which is the measure of arbitrage risk in this study. Importantly, the coefficients from this regression also indicate the optimal hedging positions.

A.4 Summary Statistics of Arbitrage Risk Measures

Table A.5 describes the summary statistics of arbitrage risk measures separately for index incumbents, additions, discretionary deletions and forced deletions.

Table A.5: Arbitrage Risk Summary Description

Variable	Mean	SD	p25	p50	p75	N
Index Incumbents						
A_1 (*10^4)	4.11	5.67	1.35	2.44	4.73	308753
A_2 (*10^4)	3.87	5.45	1.26	2.32	4.43	305430
A_0 (*10^4)	5.87	8.38	1.98	3.46	6.52	308753
E_1 (*10^4)	1.76	3.47	0.36	0.75	1.62	308753
E_2 (*10^4)	2.02	4.18	0.37	0.82	1.96	305430
R_1	0.28	0.18	0.13	0.25	0.40	308753
R_2	0.32	0.22	0.13	0.28	0.47	305430
Index Additions						
A_1 (*10^4)	5.80	6.85	1.88	3.57	6.85	652
A_2 (*10^4)	5.52	6.67	1.77	3.15	6.49	649
A_0 (*10^4)	7.95	9.87	2.55	4.56	9.32	661
E_1 (*10^4)	2.14	3.71	0.39	0.93	2.09	652
E_2 (*10^4)	2.43	4.81	0.40	0.96	2.47	649
R_1	0.24	0.15	0.13	0.22	0.33	652
R_2	0.28	0.20	0.11	0.25	0.41	649
Discretionary Index Deletions						
A_1 (*10^4)	11.35	15.55	3.33	6.52	13.56	326
A_2 (*10^4)	10.82	15.45	3.29	5.92	12.54	315
A_0 (*10^4)	14.56	19.22	4.31	7.83	16.88	326
E_1 (*10^4)	3.21	6.09	0.40	0.96	2.51	326
E_2 (*10^4)	3.80	7.60	0.27	0.83	3.34	315
R_1	0.20	0.17	0.06	0.14	0.31	326
R_2	0.22	0.22	0.04	0.15	0.35	315
Forced Index Deletions						
A_1 (*10^4)	7.06	10.80	2.25	4.03	7.33	413
A_2 (*10^4)	6.36	9.33	1.84	3.38	6.73	237
A_0 (*10^4)	8.67	14.08	2.77	4.76	8.63	414
E_1 (*10^4)	1.62	4.26	0.22	0.50	1.34	413
E_2 (*10^4)	1.81	4.44	0.20	0.53	1.53	237
R_1	0.17	0.14	0.06	0.13	0.25	413
R_2	0.20	0.20	0.06	0.14	0.32	237

This table provides a comprehensive overview of arbitrage risk and related metrics for the sample stocks. Two stock-specific arbitrage-risk measures, following the methodology of Wurgler and Zhuravskaya (2002), were constructed using daily returns over a calendar window from $[-365, -20]$ days. The first measure, A_{1i}, represents the variance of residuals from the model $R_{it} - R_{ft} = \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it}$, where R_{mt} denotes the return on CRSP's value-weighted index and R_{ft} is the risk-free rate. The second measure, A_{2i}, measures the residual variance of $R_{it} - R_{ft} = \beta_{1i}(R_{\text{sub1it}} - R_{ft}) + \beta_{2i}(R_{\text{sub2it}} - R_{ft}) + \beta_{3i}(R_{\text{sub3it}} - R_{ft}) + \varepsilon_{it}$, with R_{sub1it}, R_{sub2it}, and R_{sub3it} representing returns on three closest industry-, size-, and book-to-market-matched "substitute" stocks, matched using Fama and French 49 industry classifications. Total variance of excess returns $A_{0it} = \text{Var}(R_{it} - R_{ft})$ is estimated over the same interval as the arbitrage-risk measures. E_{1i} represents the “explained" variance $\text{Var}(R_{it} - R_{fi}) - A_{1i}$, while R_{1i} is the ratio of explained variance E_{1i} to total variance A_{0i}. E_{2i} and R_{2i} are defined analogously. All variables except ratios in the table are multiplied by (10^4) for the sake of readability. The sample period spans from January 2000 to June 2021.

A.5 Alternative Measures of Arbitrage risk

The main body of this study uses two measures of arbitrage risk that mirror those introduced by Wurgler and Zhuravskaya (2002). The two measures differ in their set of substitutes such that A_1 uses (on top of the risk-free rate) the entire market as the substitute for all stocks, while A_2 uses three stocks matched on industry and quintiles of size and book-to-market ratio. Please discuss the main text for procedural details. In a way, A_1 uses one macro-themed substitute for all the stocks, while A_2 uses micro-matched substitutes.

This section expands the arbitrage risk measures by considering alternative substitutes. Specifically, I consider the following alternative sets of substitutes: five stocks matched on industry and quintiles of size and book-to-market ratio ($A2-0-5$), en stocks matched on industry and quintiles of size and book-to-market ratio ($A2-0-10$), the market along with three stocks matched on industry and quintiles of size and book-to-market ratio ($A3-1-3$), and Fama and French (1992) three factors plus three stocks matched on industry and quintiles of size and book-to-market ratio ($A3-3-3$).

The naming convention is such that $Ai-j-k$ refers to the variance of residuals of stocks' excess returns after the optimal hedge with j macro factors (the market if $j=1$ and Fama and French (1992) three factors if $j=3$) and k individual stocks matched on industry and quintiles of size and book-to-market ratio with the stock under study. If $i=1$, this arbitrage risk uses only macro substitutes; if $i=2$, it uses only micro substitutes; and if $i=3$, it employs a mix of macro and micro substitutes. Using this naming convention, the measures A_1 and A_2 in the main text would have been called $A1-1-0$ and $A2-0-3$.

Table A.6 describes the distribution of $A2-0-5$ and $A2-0-10$, while Table A.7 presents the distribution for $A3-1-3$ and $A3-3-3$. The regression results from Table 1.4 are then replicated for these alternative arbitrage risk measures in Tables A.8 and A.9.

As readers can verify, these alternative measures only strengthen the findings presented in the main text.

Table A.6: Alternative Micro Arbitrage Risk Measures-Summary Description

Variable	Mean	SD	p25	p50	p75	N
Index Incumbents						
$A2-0-5$ (*10^4)	3.61	5.13	1.16	2.16	4.13	305315
$A2-0-10$ (*10^4)	3.21	4.71	1.01	1.90	3.66	274279
$E2-0-5$ (*10^4)	2.28	4.52	0.47	0.98	2.25	305315
$E2-0-10$ (*10^4)	2.65	5.00	0.61	1.21	2.65	274279
$R2-0-5$	0.36	0.22	0.17	0.33	0.53	305315
$R2-0-10$	0.42	0.22	0.24	0.41	0.60	274279
Index Additions						
$A2-0-5$ (*10^4)	5.15	6.36	1.62	3.01	5.95	649
$A2-0-10$ (*10^4)	4.24	5.20	1.45	2.65	5.04	558
$E2-0-5$ (*10^4)	2.80	5.15	0.53	1.23	2.88	649
$E2-0-10$ (*10^4)	3.05	5.05	0.70	1.44	3.35	558
$R2-0-5$	0.33	0.21	0.15	0.31	0.46	649
$R2-0-10$	0.39	0.21	0.23	0.38	0.54	558
Discretionary Index Deletions						
$A2-0-5$ (*10^4)	10.37	15.10	3.14	5.57	11.98	315
$A2-0-10$ (*10^4)	9.42	13.35	2.68	5.12	10.18	266
$E2-0-5$ (*10^4)	4.25	8.08	0.39	1.20	4.00	315
$E2-0-10$ (*10^4)	4.78	9.09	0.66	1.60	5.47	266
$R2-0-5$	0.25	0.23	0.06	0.19	0.43	315
$R2-0-10$	0.30	0.25	0.10	0.25	0.50	266
Forced Index Deletions						
$A2-0-5$ (*10^4)	6.12	9.16	1.74	3.24	6.55	237
$A2-0-10$ (*10^4)	5.39	7.75	1.53	2.92	6.17	209
$E2-0-5$ (*10^4)	2.05	4.64	0.31	0.77	1.70	237
$E2-0-10$ (*10^4)	2.44	5.30	0.40	1.03	2.11	209
$R2-0-5$	0.25	0.21	0.09	0.19	0.36	237
$R2-0-10$	0.30	0.22	0.13	0.24	0.44	209

This table provides a comprehensive overview of arbitrage risk and related metrics for the sample stocks. Two stock-specific arbitrage-risk measures, following the methodology of Wurgler and Zhuravskaya (2002), were constructed using daily returns over a calendar window from $[-365, -20]$ days. The first measure, A_{1i}, represents the variance of residuals from the model $R_{it} - R_{ft} = \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it}$, where R_{mt} denotes the return on CRSP's value-weighted index and R_{ft} is the risk-free rate. The second measure, A_{2i}, measures the residual variance of $R_{it}-R_{ft} = \beta_{1i}(R_{\text{sub1it}}-R_{ft})+\beta_{2i}(R_{\text{sub2it}}-R_{ft})+\beta_{3i}(R_{\text{sub3it}}-R_{ft})+\varepsilon_{it}$, with R_{sub1it}, R_{sub2it}, and R_{sub3it} representing returns on three closest industry-, size-, and book-to-market-matched "substitute" stocks, matched using Fama and French 49 industry classifications. Total variance of excess returns $A_{0it} = \text{Var}(R_{it} - R_{ft})$ is estimated over the same interval as the arbitrage-risk measures. E_{1i} represents the "explained" variance $\text{Var}(R_{it} - R_{fi}) - A_{1i}$, while R_{1i} is the ratio of explained variance E_{1i} to total variance A_{0i}. E_{2i} and R_{2i} are defined analogously. All variables except ratios in the table are multiplied by (10^4) for the sake of readability. The sample period spans from January 2000 to June 2021.

Table A.7: Alternative Mixed Arbitrage Risk Measures-Summary Description

Variable	Mean	SD	p25	p50	p75	N
Index Incumbents						
$A3-1-3$ (*10^4)	3.50	5.00	1.11	2.06	4.01	305430
$A3-3-3$ (*10^4)	3.37	4.74	1.08	2.01	3.87	305430
$E3-1-3$ (*10^4)	2.39	4.48	0.56	1.09	2.33	305430
$E3-3-3$ (*10^4)	2.52	4.72	0.61	1.17	2.46	305430
$R3-1-3$	0.38	0.21	0.22	0.36	0.54	305430
$R3-3-3$	0.40	0.21	0.24	0.38	0.56	305430
Index Additions						
$A3-1-3$ (*10^4)	5.02	6.04	1.58	2.91	5.81	649
$A3-3-3$ (*10^4)	4.83	5.74	1.56	2.77	5.60	649
$E3-1-3$ (*10^4)	2.93	5.18	0.60	1.30	2.95	649
$E3-3-3$ (*10^4)	3.12	5.40	0.68	1.43	3.14	649
$R3-1-3$	0.34	0.19	0.19	0.33	0.47	649
$R3-3-3$	0.36	0.19	0.21	0.36	0.49	649
Discretionary Index Deletions						
$A3-1-3$ (*10^4)	10.14	14.90	2.96	5.52	11.27	315
$A3-3-3$ (*10^4)	9.73	14.24	2.87	5.29	10.99	315
$E3-1-3$ (*10^4)	4.48	8.08	0.61	1.39	4.27	315
$E3-3-3$ (*10^4)	4.89	8.79	0.71	1.53	4.75	315
$R3-1-3$	0.28	0.22	0.09	0.22	0.45	315
$R3-3-3$	0.30	0.23	0.10	0.25	0.47	315
Forced Index Deletions						
$A3-1-3$ (*10^4)	6.02	8.90	1.70	3.22	6.37	237
$A3-3-3$ (*10^4)	5.89	8.69	1.67	3.18	6.26	237
$E3-1-3$ (*10^4)	2.15	4.83	0.36	0.75	1.71	237
$E3-3-3$ (*10^4)	2.27	5.06	0.39	0.79	1.93	237
$R3-1-3$	0.25	0.20	0.09	0.21	0.37	237
$R3-3-3$	0.26	0.20	0.11	0.23	0.38	237

This table provides a comprehensive overview of arbitrage risk and related metrics for the sample stocks. Two stock-specific arbitrage-risk measures, following the methodology of Wurgler and Zhuravskaya (2002), were constructed using daily returns over a calendar window from $[-365, -20]$ days. The first measure, A_{1i}, represents the variance of residuals from the model $R_{it} - R_{ft} = \beta_i(R_{mt} - R_{ft}) + \varepsilon_{it}$, where R_{mt} denotes the return on CRSP's value-weighted index and R_{ft} is the risk-free rate. The second measure, A_{2i}, measures the residual variance of $R_{it} - R_{ft} = \beta_{1i}(R_{\text{sub1it}} - R_{ft}) + \beta_{2i}(R_{\text{sub2it}} - R_{ft}) + \beta_{3i}(R_{\text{sub3it}} - R_{ft}) + \varepsilon_{it}$, with R_{sub1it}, R_{sub2it}, and R_{sub3it} representing returns on three closest industry-, size-, and book-to-market-matched "substitute" stocks, matched using Fama and French 49 industry classifications. Total variance of excess returns $A_{0it} = \text{Var}(R_{it} - R_{ft})$ is estimated over the same interval as the arbitrage-risk measures. E_{1i} represents the “explained" variance $\text{Var}(R_{it} - R_{fi}) - A_{1i}$, while R_{1i} is the ratio of explained variance E_{1i} to total variance A_{0i}. E_{2i} and R_{2i} are defined analogously. All variables except ratios in the table are multiplied by (10^4) for the sake of readability. The sample period spans from January 2000 to June 2021.

Table A.8: MRF and Alternative Micro Arbitrage Risk Measures

	(1) CAR^{Eff}_{Ann}	(2) CAR^{Eff}_{Ann}	(3) CAR^{Eff}_{Ann}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff}_{Ann}	(6) CAR^{Eff}_{Ann}
MRF	0.151	0.340	0.105	0.087	0.363	0.064
	(0.99)	(1.45)	(0.67)	(0.57)	(1.46)	(0.39)
$A2-0-5^N$	0.001^{***}	0.002	0.001^{***}			
	(5.45)	(1.61)	(4.68)			
$MRF \times A2-0-5^N$	0.355	-0.033	0.629^{***}			
	(1.64)	(-0.10)	(4.43)			
$A2-0-10^N$				0.001^{***}	0.002^{*}	0.001^{***}
				(4.90)	(1.86)	(4.27)
$MRF \times A2-0-10^N$				0.447^{**}	-0.042	0.621^{***}
				(2.22)	(-0.13)	(4.32)
Sample	All	Pre 2000	Post 2000	All	Pre 2000	Post 2000
N	285950	78205	207744	257365	67505	189859
Adj. R-sq	0.007	0.018	0.008	0.007	0.019	0.009

This table reports the results of estimating panel regression in Equation (1.9) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio, and the arbitrage risk measures divided by their standard deviation for the ease of interpretation. Arbitrage risk measures are calculated as the variance of residuals from regressing one stock's excess return on the market's excess return over a window of [-365, -20] prior days (A_1), or on the excess return of the three closest industry-, size-, and book-to-market-matched "substitute" stocks (A_2). These substitutes are matched using quintiles of size and book-to-market, employing the 49 industry classifications from Fama and French (1997). For further details, refer to Wurgler and Zhuravskaya (2002). Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: $^{*}p < 0.10; ^{**}p < 0.05; ^{***}p < 0.01$.

A.6 Demand and Volatility

This section studies the relation between non-fundamental demand shifts and stocks' volatility. For the measure of daily volatility, I use stocks' normalized price range as defined as the following equation

$$Range_{jt} = \frac{P^{High}_{jt} - P^{Low}_{jt}}{P^{close}_{j,t-1}}, \tag{A.6}$$

Table A.9: MRF and Alternative Mixed Arbitrage Risk Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	CAR^{Eff}_{Ann}	CAR^{Eff}_{Ann}	CAR^{Eff}_{Ann}	CAR^{Eff}_{Ann}	CAR^{Eff}_{Ann}	CAR^{Eff}_{Ann}
MRF	0.153	0.289	0.097	0.144	0.291	0.089
	(1.01)	(1.23)	(0.61)	(0.95)	(1.23)	(0.57)
$A3-1-3^N$	0.002***	0.003**	0.001***			
	(5.70)	(2.13)	(4.80)			
$MRF \times A3-1-3^N$	0.347*	0.029	0.606***			
	(1.66)	(0.09)	(4.45)			
$A3-3-3^N$				0.002***	0.003**	0.001***
				(5.77)	(2.02)	(4.93)
$MRF \times A3-3-3^N$				0.350*	0.027	0.604***
				(1.72)	(0.09)	(5.09)
Sample	All	Pre 2000	Post 2000	All	Pre 2000	Post 2000
N	286060	78208	207851	286060	78208	207851
Adj. R-sq	0.007	0.018	0.008	0.007	0.018	0.008

This table reports the results of estimating panel regression in Equation (1.9) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include MRF_t^j, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio, and the arbitrage risk measures divided by their standard deviation for the ease of interpretation. Arbitrage risk measures are calculated as the variance of residuals from regressing one stock's excess return on the market's excess return over a window of [-365, -20] prior days (A_1), or on the excess return of the three closest industry-, size-, and book-to-market-matched "substitute" stocks (A_2). These substitutes are matched using quintiles of size and book-to-market, employing the 49 industry classifications from Fama and French (1997). For further details, refer to Wurgler and Zhuravskaya (2002). Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: $^*p < 0.10; ^{**}p < 0.05; ^{***}p < 0.01$.

that is the difference between high and low prices normalized by the previous day's closing price.

The estimation results are reported in Table A.10. The coefficient estimates of $abs(MRF)$ are around 0.4, regardless of the controls and fixed effects. The coefficient changes only marginally when the average of the dependent variable in the previous month is added to the regressors. In terms of economic significance, given that the standard deviations of MRF and $Range$ are approximately 0.083% and 2.4%, respectively, a one standard

Table A.10: MRF and Volatility

	(1)	(2)	(3)	(4)
	$Range_t$	$Range_t$	$Range_t$	$Range_t$
$abs(MRF)$	0.384***			
	(5.73)			
$MRF > 0{=}0 \times abs(MRF)$		-0.253		
		(-1.45)		
$MRF > 0{=}1 \times abs(MRF)$		0.317***		
		(2.67)		
$Q - AvgRange_M{=}1 \times abs(MRF)$			-0.237***	
			(-2.91)	
$Q - AvgRange_M{=}2 \times abs(MRF)$			-0.304***	
			(-3.40)	
$Q - AvgRange_M{=}3 \times abs(MRF)$			0.141*	
			(1.95)	
$Q - AvgRange_M{=}4 \times abs(MRF)$			1.954***	
			(16.55)	
era=0 × $abs(MRF)$				0.590***
				(12.53)
era=1 × $abs(MRF)$				0.212
				(1.07)
era=2 × $abs(MRF)$				0.136
				(1.01)
Controls	Y	Y	Y	Y
FE	Y & S	Y & S	Y & S	Y & S
N	289014	208977	289014	289014
Adj. R-sq	0.412	0.626	0.413	0.399

This table reports the results of estimating panel regression in Equation (1.7) for index incumbents when the dependent variable is daily volatility, measured as the daily price range (high - low) over the previous day's closing price. The sample includes all incumbent observations in the reconstitution days in the sample period (1990-01, 2021-06). The independent variable is the absolute value of MRF_t^i, the surprise dollar amount of money flowed into stock i at the reconstitution day t (proportional to the previous day's market value) as a result of the mechanical rebalancing of index funds. Control variables include $\log MV$ (the logarithm of proprietary total market value), IWF (proprietary float factor), and β^{CAPM} (loading on the market in the one-factor market model). T-statistics based on standard errors double-clustered by stock and day are in parentheses. Significance levels are marked as: $^{*}p < 0.10; ^{**}p < 0.05; ^{***}p < 0.01$.

deviation increase in MRF results in an increase in the daily stock turnover that is 1.3% of this variable's sample standard deviation.

Regression (3) indicates that the increase in daily volatility of stocks is larger for positive demand shifts. This complements the earlier discussion on arbitrage risk. When the demand shift is positive (i.e., indexers need to buy a stock), arbitrageurs must sell to them, possibly through short selling, which incurs costs. The arbitrageur, facing short selling costs, is willing to sell to the indexers only at a higher price to compensate for the additional cost of short selling. This finding aligns with the theoretical prediction in Gromb and Vayanos (2010) that in the presence of short selling costs, the price effect of demand shifts can be asymmetric.

Regressions (4) and (8) provide additional insights by demonstrating that the average increase in volatility resulting from index funds' rebalancing is not significantly different for stocks that are unconditionally more or less volatile. In other words, the impact on volatility due to index funds' rebalancing does not depend on a stock's inherent level of volatility, measured as the quartiles of daily volatility in the past month.

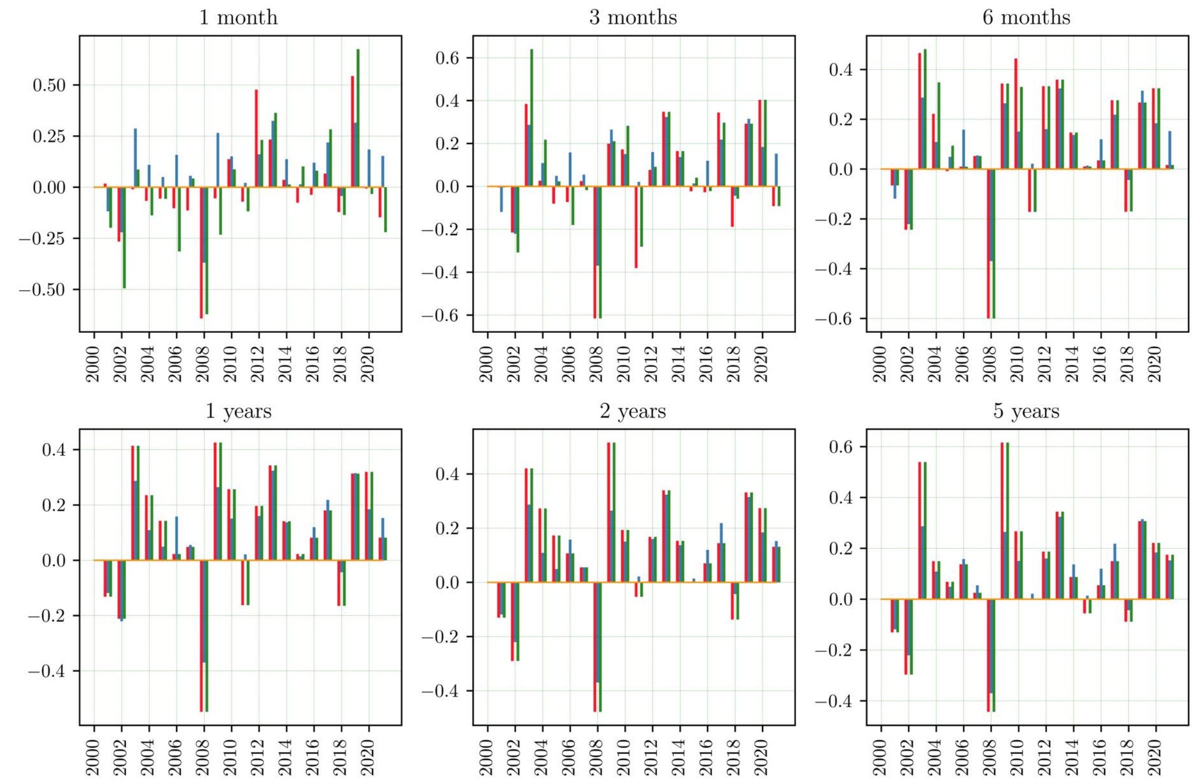
The findings in this section also contribute to the understanding of the *excess volatility puzzle*[1]. Basak and Pavlova (2013) employ a theoretical framework demonstrating how index funds can amplify volatility within their benchmark stocks and across the broader stock market. Among the empirical works, Ben-David et al. (2018) illustrate that ETFs, primarily passive index trackers, can increase the volatility of their portfolio stocks through a liquidity channel. They show that ETFs may elevate the non-fundamental volatility of securities in their baskets by transmitting liquidity shocks to underlying securities via arbitrage activities. The findings in this section further support the notion that demand shifts unrelated to fundamentals can amplify stock price volatility. However, unlike the liquidity channel discussed in Ben-David et al. (2018), the volatility increase here arises from direct trades by indexers in an imperfectly efficient market rather than through arbitrage between ETF prices and underlying basket prices.

[1] The excess volatility puzzle (Shiller, 1981) is the observation in financial markets that asset price volatility exceeds what can be explained by changes in underlying fundamentals. Specifically for stocks, it refers to the phenomenon where stock prices fluctuate more than can be justified by changes in expected dividends or earnings of firms.

A.7 Long-term Returns

This section investigates the returns of S&P additions and deletions over longer horizons. The purpose is to determine whether adding a company to the index is a form of reward for its past exemplary performance or a reflection of its future potential for success. Similarly, it aims to discern whether removal from the index is a penalty for substandard past performance or a result of inferior prospects. The findings indicate that S&P inclusion and exclusion decisions are more closely linked to a company's past performance and stock returns, thereby doubting the hypothesis that these decisions are based on insider information or superior analytical power. I consider only discretionary deletions in this

Figure A.1: Calendar-time portfolio returns over time



The figure shows the annual returns of the calendar-time portfolios constructed by recently added (green) and recently deleted (red) stocks. The annual return of the S&P 500 total return index (blue) is added as a benchmark. The holding period of the stocks in the portfolios after the event is mentioned above the figures. The sample period is (2000-01, 2021-06).

section since forced deletions don't exist after the event.

To examine long-term returns, I employed calendar-time portfolios due to their ability to account for cross-sectional correlation among constituent stocks within the portfolio, as opposed to CARs. In addition, these portfolios offer a practical investment strategy. I created two groups of equal-weighted portfolios based on recently added and deleted stocks using the following approach: for each holding period of N days, the corresponding portfolios are comprised of stocks added to or removed from the index during the preceding N days. I repeated this strategy for multiple holding periods to identify trends. These portfolios were constructed at the outset of the sample period and rebalanced daily until the end of it. Dividends were reinvested in the corresponding stocks. If a portfolio was empty in a day, its value was kept still until the next time a stock was added to it.

Figure A.1 shows each year's annual return of constructed portfolios. S&P 500 total return index is also added to the figures for better comparison. While there are many variations between the added and the deleted portfolios for short holding periods, there is virtually no difference between their returns in longer horizons. Furthermore, in most of the years, the direction of their returns agrees with that of the index.

Table A.11 presents a comparison of portfolio returns in a different manner. First, long-horizon portfolios that held assets in them for over a year generated a larger average return than the index. However, their sharp ratios were much lower than that of the index. Second, none of the portfolios, including the index, produced a positive *alpha*. In particular, both portfolios yielded a negative and significant *alpha* for short holding periods. These findings provide evidence against the information hypothesis. Third, Chan et al. (2013) found positive and significant *alpha*s on similar equal-weighted calendar-time portfolios using a sample that predates the one used in this study. Therefore, it appears that even if S&P 500 changes were informative in the past, they are no longer so.

Table A.11: Calendar-time portfolio return characteristics

Portfolio	Holding Period	Mean	STD	Sharpe Ratio	Num Stocks	α	β_{MKT}	β_{SMB}	β_{HML}	β_{UMD}	R^2
S&P 500	-	0.67	4.35	0.12	501.46	0.01	0.98***	-0.17***	0.02***	-0.02***	99.51
Additions	1 month	-0.42	8.06	-0.07	2.01	-1.22***	0.94***	0.42***	-0.01	0.09	33.51
Additions	3 months	0.35	7.14	0.03	6.00	-0.57*	1.12***	0.46***	-0.09	0.14**	59.24
Additions	6 months	0.60	6.38	0.07	11.96	-0.28	1.15***	0.32***	-0.16**	0.10**	73.33
Additions	1 year	0.57	6.13	0.07	23.95	-0.25	1.10***	0.34***	-0.28***	-0.05	83.07
Additions	2 years	0.69	6.27	0.09	46.62	-0.12	1.12***	0.32***	-0.26***	-0.15***	86.65
Additions	5 years	0.73	6.52	0.09	105.52	-0.12	1.15***	0.39***	-0.19***	-0.21***	89.24
Deletions	1 month	-0.30	6.18	-0.05	0.88	-0.84**	0.52***	0.40***	-0.00	0.01	23.23
Deletions	3 month	0.12	6.49	-0.00	2.59	-0.68**	0.97***	0.41***	-0.17**	0.06	56.62
Deletions	6 month	0.59	6.31	0.07	5.08	-0.28	1.12***	0.36***	-0.18***	0.09**	73.39
Deletions	1 year	0.61	6.09	0.08	10.02	-0.21	1.09***	0.36***	-0.29***	-0.06*	83.58
Deletions	2 years	0.73	6.23	0.10	18.95	-0.08	1.11***	0.35***	-0.27***	-0.15***	87.13
Deletions	5 years	0.76	6.48	0.10	40.23	-0.08	1.13***	0.41***	-0.20***	-0.22***	89.72

The table compares the monthly returns of the calendar-time portfolio constructed by recently added or deleted stocks. The second column indicates how long each stock is held in the portfolio after the event. Each stock's dividends are reinvested in it, and the first row shows similar measures for the S&P 500 total return index as a benchmark. Columns (3) and (4)) show, respectively, the arithmetic mean and standard deviation of monthly returns. Column (5) measures the Sharpe ratio, defined as the average of monthly excess returns over their standard deviation. Column (6) shows the average number of stocks in the portfolios. Columns (7) to (12) show the results of the regression of monthly returns based on Carhart (1997) four-factor model. The sample period is (2000-01, 2021-06). Significance levels are marked as: *p < 0.10; **p < 0.05; ***p < 0.01.

A.8 Fundamental Analysis

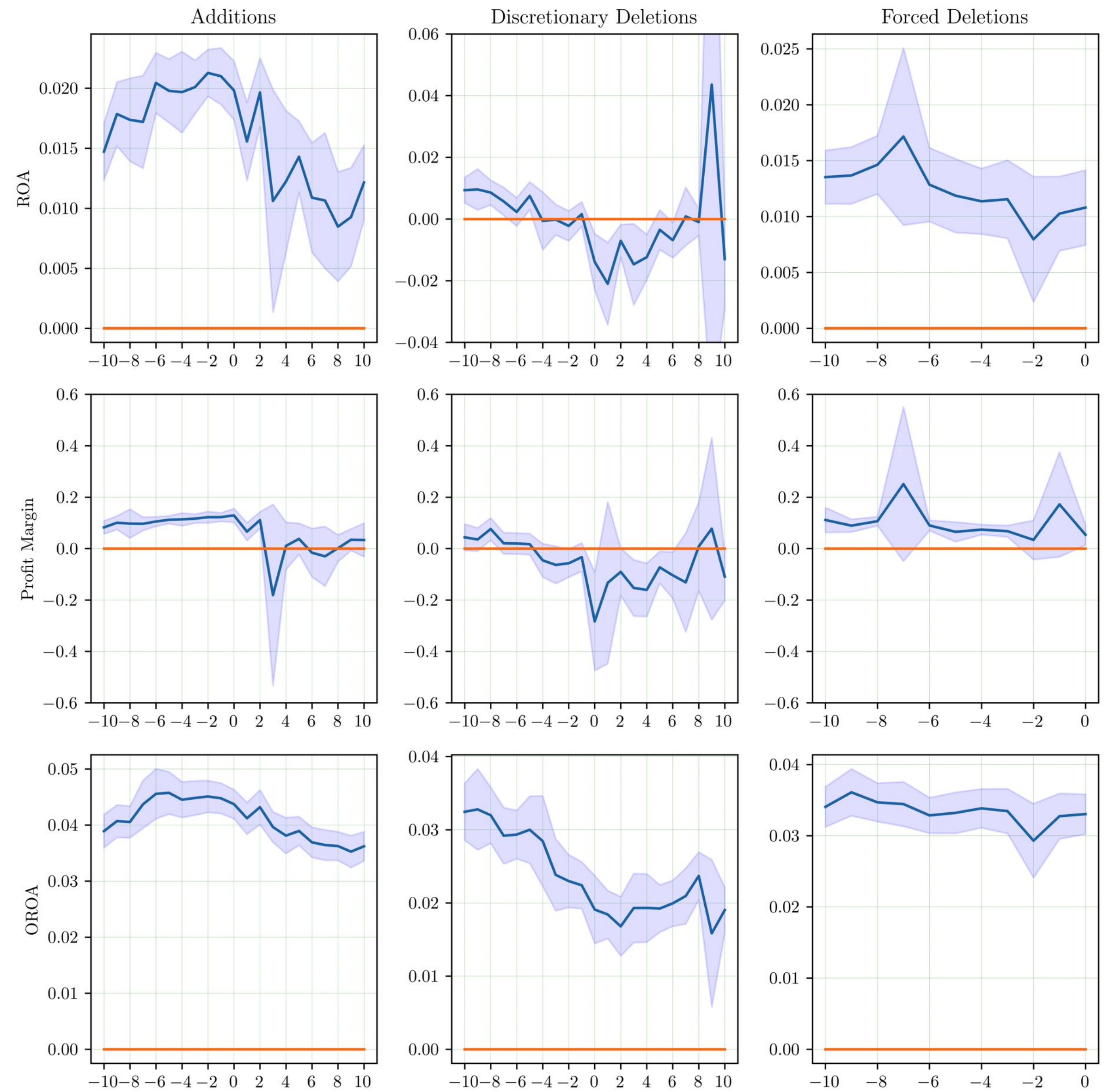
In this section, I analyze the fundamental characteristics of the stocks added and deleted from the S&P 500 index using an event-study approach. The study uses financial ratios to ensure comparability, with quarter zero representing the last fiscal quarter for which balance sheet data was available at the time of the index announcement. To measure profitability and efficiency, I consider three key metrics: Return on Assets (ROA), Profit Margin (net income/sales), and Operating ROA (operating income/assets), where operating income is defined as operating income before depreciation, amortization, and taxes, plus interest income.

The event study results shown in Figure A.2 reveal that, for all the suggested measures, the S&P 500 additions are at their highest around the time of the index announcement. This result suggests that the index committee selected firms for inclusion when they had recently experienced several quarters of growth, leading to their best recent stance, as reflected in their financial ratios. However, this good performance is not necessarily sustainable and is subject to mean reversion, as all the measures decline right after the inclusion.

Index deletion exhibits an inverse pattern. These firms are deleted from the index right after they report their first negative net income, usually after several quarters of decreasing net income starting approximately two years before the deletion event. Decreasing operating ROA in the third row confirms that the negative net income for these firms is not just a financial figure due to higher interest paid, which is coming from lower operating efficiency. The figure also shows that a similar mean reversion mechanism also works for these firms, meaning that after they are deleted from the index at their lowest-ever instance, they start to recover. In five years, most measures are already near their previous average amounts.

On the contrary, the pattern of S&P 500 discretionary deletions is characterized by

Figure A.2: Fundamental trends in index additions and deletions



The figure compares fundamentals in index additions and deletions in a quarterly event study framework. Quarter zero represents the last fiscal quarter for which balance sheet data was available at the index announcement. Confidence intervals are shown in light shadow around the means. Fundamental measures are calculated as ROA = net income/ assets, Profit Margin = net income/ Sales, and OROA = operating income/ assets. The sample period is (2000-01, 2021-06). 95% confidence intervals are shown in light shadow around the means.

firms being excluded from the index after reporting a negative net income, preceded by a trend of declining net income over multiple quarters. This decline in net income is not solely attributable to higher interest payments but also reflects decreased operating efficiency, as indicated by declining operating return on assets (ROA). After their removal from the index, these firms tend to experience a mean reversion effect, and most of their measures start to recover, approaching their previous levels. The measures for forced deletions, on the other hand, appear to be relatively unchanged, reassuring that these firms' removal from the index was likely a result of external factors rather than their financial performance.

A.9 Linearity of price impact to demand shifts

This section explores the linearity of price impact in response to demand shifts. Verifying this linearity is crucial because the counterfactual scenarios in Section 1.4.2 depend on price impact estimates for incumbents, which are derived from relatively small demand shifts, but these estimates are then applied to predict actual additions and deletions involving much larger demand shifts.

To assess linearity, I add the squared value of MRF to the baseline regression specifications in (1.7) when the outcome variable is CAR_{Ann}^{Eff}–the total abnormal return from the announcement to the effective day. The results of this estimation are reported in Table A.12. Different specifications also test linearity separately over various periods, both pre- and post-2000, as well as with and without firm controls. In none of the specifications does the coefficient of the squared value of MRF appear significant, which confirms that the price impact in response to demand shifts is linear with respect to the size of these shifts.

Table A.12: MRF and Price Impact-Tests of Linearity

Dependent Variable	(1) CAR^{Eff}_{Ann}	(2) CAR^{Eff}_{Ann}	(3) CAR^{Eff}_{Ann}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff}_{Ann}
MRF	0.416***	0.127	0.579***	0.544**	0.387***
	(3.49)	(0.65)	(2.67)	(2.46)	(3.55)
MRF^2	-0.295	-48.236	-2.004	-1.443	-0.224
	(-0.11)	(-1.32)	(-0.74)	(-0.51)	(-0.09)
Adj. R-sq	0.006	0.017	0.008	0.016	0.004
N	288997	80033	208963	208963	308726
Sample	All	Pre-2000	Post-2000	Post-2000	All
Controls	Yes	Y	Yes	Yes	NO
Time-FE	Year	Year	Year	Day	Year
Stock-FE	Yes	Yes	Yes	Yes	Yes
Clusters	S	S	S	S	S

This table presents the results of estimating the panel regression specified in Equation (1.7) for index incumbents. The full sample comprises all incumbent observations on reconstitution days throughout the entire sample period (1990-01 to 2021-06). Results are then replicated separately for three sub-samples: Early (1990-1999), Mid (2000-2010), and Late (2011-2021). The bottom rows assess the statistical significance of the differences between sub-sample estimates. The independent variable is MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio. All regressions include stock fixed effects, and year fixed effects. All abnormal returns are calculated using coefficients from a four-factor Carhart (1997) model incorporating a moving window of [-365, -20] days prior to the day under study, conditional on having 100 observations. T-statistics based on standard errors clustered by stocks are reported in parentheses. Significance levels are denoted as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

A.10 Alternative specifications for industry-matched firms

Table A.13 replicates Table 1.8 using two-digit SIC codes to match incumbents' industries (instead of three-digit codes as used in the main text). The results portray a similar picture for this alternative specification. Tables A.14 and A.15 estimate the main demand effect of interest in this pape–the impact of MRF on the cumulative abnormal returns of incumbents from the announcement to the effective day of index reconstitutions–in addition to the informational effect on firms within the same industry of additions and deletions, following the tradition of Cai (2007). These tables demonstrate that the demand effect is not driven by and is independent of this informational channel.

Table A.13: Index effect and industry-matched incumbents

	(1) CAR^{Eff}_{Ann}	(2) CAR^{Eff}_{Ann}	(3) CAR^{Eff}_{Ann}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff}_{Ann}
SameInd2Add	0.001***				
	(3.00)				
Post-2000=0 × SameInd2Add		0.003***			
		(3.27)			
Post-2000=1 × SameInd2Add		0.001			
		(1.36)			
SameInd2Del			0.000		
			(0.51)		
Post-2000=0 × SameInd2Del				0.002*	
				(1.72)	
Post-2000=1 × SameInd2Del				-0.000	
				(-0.39)	
Post-2000=0 × SameSIC2Del × Discretionary					0.001
					(0.44)
Post-2000=1 × SameSIC2Del × Discretionary					-0.000
					(-0.26)
Controls	Y	Y	Y	Y	Y
FE	Y & S	Y & S	Y & S	Y & S	Y & S
N	288997	288997	288997	288997	288997
Adj. R-sq	0.006	0.006	0.006	0.006	0.006

This table reports the results of estimating panel regression in Equation (1.7) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include dummy variables if the incumbents are in the same two digit SIC industry with additions or deletions. Discretionary is a dummy variable that is equal to one if there is a discretionary deletion in respective reconstitution and zero otherwise. Post-2000 is dummy variable indicating year 2000 or afterwards. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.14: Index effect and industry-matched incumbents

	(1) CAR_{Ann}^{Eff}	(2) CAR_{Ann}^{Eff}	(3) CAR_{Ann}^{Eff}	(4) CAR_{Ann}^{Eff}	(5) CAR_{Ann}^{Eff}
MRF	0.413***	0.416***	0.411***	0.413***	0.409***
	(3.59)	(3.62)	(3.57)	(3.58)	(3.56)
SameSIC3Add	0.001**				
	(2.52)				
Post-2000=0 × SameSIC3Add		0.003**			
		(2.41)			
Post-2000=1 × SameSIC3Add		0.001			
		(1.33)			
SameSIC3Del			-0.001		
			(-1.61)		
Post-2000=0 × SameSIC3Del				0.003	
				(1.47)	
Post-2000=1 × SameSIC3Del				-0.003***	
				(-2.70)	
Post-2000=0 × SameSIC3Del × Discretionary					-0.007*
					(-1.94)
Post-2000=1 × SameSIC3Del × Discretionary					0.001
					(0.71)
Controls	Y	Y	Y	Y	Y
FE	Y & S	Y & S	Y & S	Y & S	Y & S
N	288997	288997	288997	288997	288997
Adj. R-sq	0.006	0.006	0.006	0.006	0.006

This table reports the results of estimating panel regression in Equation (1.7) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include MRF_t^j, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio, and dummy variables if the incumbents are in the same three digit SIC industry with additions or deletions. Discretionary is a dummy variable that is equal to one if there is a discretionary deletion in respective reconstitution and zero otherwise. Post-2000 is dummy variable indicating year 2000 or afterwards. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: $^*p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table A.15: Index effect and industry-matched incumbents

	(1) CAR^{Eff}_{Ann}	(2) CAR^{Eff}_{Ann}	(3) CAR^{Eff}_{Ann}	(4) CAR^{Eff}_{Ann}	(5) CAR^{Eff}_{Ann}
MRF	0.414***	0.418***	0.411***	0.411***	0.412***
	(3.59)	(3.64)	(3.57)	(3.57)	(3.57)
SameInd2Add	0.001***				
	(3.05)				
Post-2000=0 × SameInd2Add		0.003***			
		(3.36)			
Post-2000=1 × SameInd2Add		0.001			
		(1.36)			
SameInd2Del			0.000		
			(0.52)		
Post-2000=0 × SameInd2Del				0.002*	
				(1.73)	
Post-2000=1 × SameInd2Del				-0.000	
				(-0.38)	
Post-2000=0 × SameSIC2Del × Discretionary					0.001
					(0.46)
Post-2000=1 × SameSIC2Del × Discretionary					-0.000
					(-0.28)
Controls	Y	Y	Y	Y	Y
FE	Y & S	Y & S	Y & S	Y & S	Y & S
N	288997	288997	288997	288997	288997
Adj. R-sq	0.006	0.006	0.006	0.006	0.006

This table reports the results of estimating panel regression in Equation (1.7) on index incumbents. The dependent variable in all regressions is the cumulative abnormal returns of index incumbents from the announcement day to the effective day of index reconstitutions. The independent variables include MRF^j_t, representing the dollar amount of money flowing into or withdrawn from stock j on reconstitution day t (proportional to the previous day's market value) due to mechanical weight changes in the index portfolio, and dummy variables if the incumbents are in the same two digit SIC industry with additions or deletions. Discretionary is a dummy variable that is equal to one if there is a discretionary deletion in respective reconstitution and zero otherwise. Post-2000 is dummy variable indicating year 2000 or afterwards. Control variables include $\log MV$ (the logarithm of total market value), IWF (proprietary float factor), β^{CAPM} (loading on the market in the one-factor market model, estimated over a one-year window ending 20 days before the observation day), and liquidity (the average daily bid-ask spread of the stock over a one-month window ending 20 trading days before the observation day). All regressions include control variables, stock fixed effects, and year fixed effects. T-statistics based on standard errors clustered by stocks are reported in parentheses. The sample period is (1990-01, 2021-06). Significance levels are marked as: $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}p < 0.01$.

References

Afego, Pyemo N., 2017, Effects of changes in stock index compositions: A literature survey, *International Review of Financial Analysis* 52, 228–239.

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Basak, Suleyman, and Anna Pavlova, 2013, Asset prices and institutional investors, *American Economic Review* 103, 1728–1758.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2018, Do etfs increase volatility?, *Journal of Finance* 73, 2471–2535.

Beneish, M.D., and R.E. Whaley, 1996, An anatomy of the "s&p game": The effects of changing the rules, *Journal of Finance* 51, 1909–1930, cited By 188.

Bennett, Benjamin, Rene M. Stulz, and Zexi Wang, 2023, Keeping up with the joneses and the real effects of s&p 500 inclusion, *SSRN Electronic Journal* .

Cai, Jie, 2007, What's in the news? information content of s&p 500 additions, *Financial Management* 36, 113–124.

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

Chan, Kalok, Hung Wan Kot, and Gordon YN Tang, 2013, A comprehensive long-term analysis of s&p 500 index additions and deletions, *Journal of Banking & Finance* 37, 4920–4930.

Chang, Yen-Cheng, Harrison Hong, and Inessa Liskovich, 2014, Regression Discontinuity and the Price Effects of Stock Market Indexing, *Review of Financial Studies* 28, 212–246.

Chaudhary, Manav, Zhiyu Fu, and Jian Li, 2022, Corporate bond multipliers: Substitutes matter, *SSRN Electronic Journal* .

Chen, Honghui, Gregory Noronha, and Vijay Singal, 2004, The price response to s&p 500 index additions and deletions: Evidence of asymmetry and a new explanation, *Journal of Finance* 59, 1901–1930.

Chinco, Alex, and Marco Sammon, 2024, The passive ownership share is double what you think it is, *Journal of Financial Economics* 157, 103860.

Danbolt, Jo, Ian Hirst, and Edward Jones, 2018, Gaming the ftse 100 index, *British Accounting Review* 50, 364–378.

Denis, Diane K., John J. McConnell, Alexei V. Ovtchinnikov, and Yun Yu, 2003, S&p 500 index additions and earnings expectations, *Journal of Finance* 58, 1821–1840.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427–465.

Fama, Eugene F., and Kenneth R. French, 1997, Industry costs of equity, *Journal of Financial Economics* 43, 153–193.

Fernandes, Marcelo, and João Mergulhão, 2016, Anticipatory effects in the ftse 100 index revisions, *Journal of Empirical Finance* 37, 79–90.

Fuchs, William Martin, Satoshi Fukuda, and Daniel Neuhann, 2023, Demand-system asset pricing: Theoretical foundations, *SSRN Electronic Journal* .

Gabaix, Xavier, and Ralph S. J. Koijen, 2022, In search of the origins of financial fluctuations: The inelastic markets hypothesis, *SSRN Electronic Journal* .

Gabaix, Xavier, Ralph S. J. Koijen, Federico Mainardi, Sangmin Oh, and Motohiro Yogo, 2023, Asset demand of u.s. households, *SSRN Electronic Journal* .

Greenwood, Robin, 2005, Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics* 75, 607–649.

Greenwood, Robin M., and Marco Sammon, 2024, The disappearing index effect, *Journal of Finance* Forthcoming.

Gromb, Denis, and Dimitri Vayanos, 2010, Limits of arbitrage, *Annu. Rev. Financ. Econ.* 2, 251–275.

Haddad, Valentin, Paul Huebner, and Erik Loualiche, 2021, How competitive is the stock market? theory, evidence from portfolios, and implications for the rise of passive investing, *SSRN Electronic Journal* 3821263.

Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the s&p 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41, 815–829.

Hartzmark, Samuel M, and David H Solomon, 2022, Predictable price pressure, *SSRN Electronic Journal* .

Heath, Davidson, Daniele Macciocchi, Roni Michaely, and Matthew C. Ringgenberg, 2020, Do index funds monitor?, *Review of Financial Studies* .

Honkanen, Pekka, and Daniel Schmidt, 2022, Learning from noise? price and liquidity spillovers around mutual fund fire sales, *Review of Asset Pricing Studies* 12, 593–637.

Kaul, Aditya, Vikas Mehrotra, and Randall Morck, 2000, Demand curves for stocks do slope down: New evidence from an index weights adjustment, *Journal of Finance* 55, 893–912.

Koijen, Ralph S J, Robert J Richmond, and Motohiro Yogo, 2023, Which Investors Matter for Equity Valuations and Expected Returns?, *Review of Economic Studies* 91, 2387–2424.

Koijen, Ralph SJ, and Motohiro Yogo, 2019, A demand system approach to asset pricing, *Journal of Political Economy* 127, 1475–1515.

Kumar, Rajnish, Edward R. Lawrence, Arun Prakash, and Iván M. Rodríguez, 2023, Additions to and deletions from the s&p 500 index: A resolution to the asymmetric price response puzzle, *Journal of Banking and Finance* 154, 106976.

Li, Jiacui, and Zihan Lin, 2022, Price multipliers are larger at more aggregate levels, *SSRN Electronic Journal* 4038664.

Lou, Dong, 2012, A Flow-Based Explanation for Return Predictability, *The Review of Financial Studies* 25, 3457–3489.

Lynch, Anthony W, and Richard R Mendenhall, 1997, New evidence on stock price effects associated with changes in the s&p 500 index, *Journal of Business* 70, 351–383.

Mase, Bryan, 2007, The impact of changes in the ftse 100 index, *Financial Review* 42, 461–484.

Pandolfi, Lorenzo, and Tomas Williams, 2019, Capital flows and sovereign debt markets: Evidence from index rebalancings, *Journal of Financial Economics* 132, 384–403.

Patel, Nimesh, and Ivo Welch, 2017, Extended stock returns in response to s&p 500 index changes, *Review of Asset Pricing Studies* 7, 172–208.

Pavlova, Anna, and Taisiya Sikorskaya, 2022, Benchmarking Intensity, *Review of Financial Studies* 36, 859–903.

Petajisto, Antti, 2009, Why do demand curves for stocks slope down?, *Journal of Financial and Quantitative Analysis* 44, 1013–1044.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135–1151.

Schmickler, Simon, 2020, Asset fire sales or assets on fire?, *SSRN Electronic Journal* .

Shiller, Robert J, 1981, Do stock prices move too much to be justified by subsequent changes in dividends?, *American Economic Review* .

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41, 579–590.

Shleifer, Andrei, and Robert W Vishny, 1997, The limits of arbitrage, *Journal of finance* 52, 35–55.

Sikorskaya, Taisiya, 2023, Institutional investors, securities lending, and short-selling constraints, *SSRN Electronic Journal* 4627075.

Soe, Aye M., and Srikant Dash, 2008, The shrinking index effect: A global perspective, *Standard & Poor's Publication* .

Vayanos, Dimitri, and Paul Woolley, 2013, An institutional theory of momentum and reversal, *Review of Financial Studies* 26, 1087–1145.

Vijh, Anand M., and Jiawei (Brooke) Wang, 2022, Negative returns on addition to the s&p 500 index and positive returns on deletion? new evidence on the attractiveness of s&p 500 versus s&p 400 indexes, *Financial Management* 51, 1127–1164.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, *Journal of Business* 75, 583–608.

Chapter 2

Corporate Debt Structure Around the World

joint with Stefano Rossi and Lorenzo Bretscher

Abstract

Combining different datasets for the first time, we reconstruct the debt structure of 10,136 firms operating across 52 countries from 2003 to 2021, which comprise 124,101 corporate bond issues held by 63,323 bondholders and 85,817 bank loans and credit lines extended by 10,154 banks. Contrary to prior literature, total debt ownership is most concentrated in civil law countries, and most dispersed in common law countries. This result is driven by dispersion of junior arm's length debt. There is considerable heterogeneity: In countries with stronger investor protection the debt structure features dispersed debt ownership, with dispersed arm's length borrowing coexisting with concentrated bank borrowing. In countries with weaker investor protection, firms mitigate the unfavorable institutional environment by borrowing at shorter maturity, and by borrowing in USD-denominated (foreign) debt. These patterns are strongest for small-to-medium size firms, whereas the largest firms have dispersed debt ownership by international investors almost irrespective of their country of incorporation. The considerable heterogeneity that we uncover along debt types and firm size thus supports the importance of considering at the same time different sources of debt, including bank loans, credit lines, and arm's length debt, to obtain a full picture of firms' debt structures around the world. Our results broadly support legal origin and financial contracting theories of corporate debt structure.

Keywords: Debt Structure, Debt Ownership Concentration, Law and Finance, Investor Protection, Creditor Rights.

2.1 Introduction

Understanding the determinants and consequences of different sources of external financing is a first order question in corporate finance. While there has been a lot of work to date on corporate equity ownership and control around the world (e.g., La Porta et al. (1998), La Porta et al. (1999), Aminadav and Papaioannou (2020)), corporate debt structures have received less attention. Given that debt represents about 85% of the external financing of firms (Beck et al. (2008)) and debt structures determine the allocation of power and assets in financial distress and bankruptcy, it is important to study corporate debt structures around the world.

Debt structures are complex and multidimensional. For instance, some firms borrow from a single bank, using loans or credit lines; others borrow from multiple banks; others issue bonds to arm's length investors; and others borrow using a combination of debt classes from a multiplicity of borrowers. A sizable literature has studied various individual aspects of firms' debt structures across countries, but the literature has yet to reach a consensus, some times even on basic correlations, due to a number of empirical challenges.

The first challenge is that no single database covers all possible debt contracts and reports the identity of all lenders. Datasets covering bank debt contracts either focus on specific debt markets, e.g., individual bank loans (Ongena and Smith (2000), Detragiache et al. (2000)), or the syndicated loan markets (Esty and Megginson (2003); Qian and Strahan (2007); Bae and Goyal (2009)), and do not cover arm's length contracts. Conversely, datasets covering bond issuances do not cover bank debt (Manconi et al. (2016), Bretscher et al. (forthcoming)). Furthermore, datasets covering different debt classes typically do not record the identity of the lenders (Colla et al. (2013); John et al. (2021)). Finally, datasets covering capital structure do not distinguish among debt classes and do not record the identity of the lenders (Demirgüç-Kunt and Maksimovic (1999), Giannetti (2003), Fan et al. (2012)).

A second challenge relates to sample size and composition. Some of these datasets cover predominantly large firms, others cover small-to-medium size firms. A third challenge is heterogeneity with respect to size and age of sample firms, as the size distribution of firms is highly skewed, and private and public firms at different stages of their life cycle differ systematically in their debt structure. Finally, unlike in the case of equity, no regulation mandates the disclosure of large blocks of bond holdings in excess of some threshold. As a result, reconstructing the full debt structure of firms is challenging, as it is possible to miss sizeable untraded or privately placed blocks of debt.

We make progress on each of these fronts by combining for the first time different databases including Amadeus, Capital IQ, Dealscan, eMaxx, and LSEG Refinitiv to reconstruct the structure of outstanding debt at the firm level, including bank loans, syndicated loans, bonds, credit lines, and by collating this information to reconstruct as large a fraction as possible of the outstanding debt ownership. We end up with a comprehensive dataset reconstructing the debt structure of 10,136 firms operating across 52 countries from 2003 to 2021, which comprises 124,101 corporate bond issues held by 63,323 bondholders and 85,817 bank loans and credit lines extended by 10,154 banks.

Debt structures allocate power across creditors and between creditors and debtors. Economic theories of creditor power start from the observation that when lenders can more easily force repayment, repossess collateral, or even gain control of the firm following non-repayment ex post, they are more willing to lend ex ante. These power theories of credit have been formalized in a one-creditor context by Townsend (1979), Aghion and Bolton (1992), and Hart and Moore (1994, 1998).

Building on this framework, there are two views of debt structure and creditor power in a context of creditor multiplicity. Under one view, high debt ownership dispersion substitutes for low investor protection. The reason is that debt ownership dispersion protects creditors against strategic default (Bolton and Scharfstein, 1996), which makes dispersion valuable at low levels of investor protection where strategic default is problematic

(Diamond, 2004). Diamond (2004) also emphasizes that when investor protection is low creditors should protect themselves by lending at short maturity.

Under another view, high debt ownership dispersion is an outcome of high investor protection. The key idea here is that small, diversified investors can only be prevalent when investor protection is high (La Porta et al., 1998, 2008). Gennaioli and Rossi (2013) present a model where concentrated debt with exclusive control rights over the liquidation/reorganization decision coexists with dispersed junior creditors without control rights. In such a model, dispersion of junior debt protects the borrower against inefficient liquidation, which makes dispersion valuable at relatively high levels of investor protection, where rescuing profitable firms ex post is valuable. Gennaioli and Rossi (2013) also emphasize how securing debt with collateral should be easier when investor protection is higher.

While the two views yield opposite predictions on the dispersion of total debt ownership, the two views are not necessarily mutually exclusive, at least as soon as one recognizes that Bolton and Scharfstein (1996) and Diamond (2004) model a single debt class of "large" creditors without considering the allocation of control rights over the liquidation/reorganization decision, whereas Gennaioli and Rossi (2013) study the allocation of control rights and allow for multiple classes of debt, but do not consider debt maturity.

In this paper, we focus on five aspects of the debt structure related to creditor power: What determines the number of creditors a firm borrows from? What determines how many classes of debt firms use? What determines the maturity of (different types of) debt contracts? What determines the allocation of security interests among creditors and debt types? What determines the currency of debt contracts? Our analysis shows that there are large differences in debt structures across legal families and yields the following results.

(i) Total debt ownership is most concentrated in civil law countries, and it is most dispersed in common law countries. There is significant heterogeneity with respect to debt types: ownership dispersion in arm's length debt is largest in common law countries, whereas, if anything, firms borrow from more banks in civil law countries. Hence, the result that total debt ownership dispersion is largest in common law countries is driven by bond ownership dispersion. This is a novel result. As we discuss below in detail, prior literature has examined only bank debt, documenting the opposite result of higher total debt ownership dispersion in civil law countries.

(ii) Firms in civil law countries borrow at shorter maturity. When distinguishing between bank debt and arm's length debt, we find that firms in common law countries issue bonds at significantly longer maturity; whereas bank loans and credit lines have very similar maturity across legal origins. As a result, the finding that maturity is shorter in civil law countries is driven by the maturity of arm's length debt.

(iii) Firms in common law countries are more likely to have their debt secured by some collateral. When distinguishing between bank debt and arm's length debt, we find that firms in common law countries are more likely to have secured bank debt, whereas arm's length debt is similarly (un-)secured in common law and civil law countries.

(iv) Firms in civil law countries are more likely to borrow in USD-denominated debt contracts. The USD-denominated proportion of both bank loans and arm's length debt is larger in French civil law countries. On the other hand, German civil law countries borrow in local currencies.

(v) Whereas total corporate leverage is similar in common law and civil law countries, firms in civil law countries are more likely to borrow from banks and less likely to issue arm's length debt than firms in common law countries.

Next, we examine the correlation between corporate control and institutional characteristics that legal origin theories emphasize in a simple, unified framework, employing multiple proxies for institutional quality to account for specific theoretical predictions. Namely, the creditor rights index (Djankov et al. (2007)) measures the rights of secured creditors in bankruptcy, and we use it accordingly to examine cross-country variation in secured debt. On the other hand, the anti-self-dealing index (Djankov et al. (2008b)) measures the legal protection available to minority investors against managerial self dealing in terms of both disclosure and enforcement, and although it is defined specifically for minority shareholders it is likely also relevant for dispersed, arm's length creditors.[1] Although the cross-country associations are not well suited to advance causality, they are informative on the characteristics of the legal and institutional environment that relate to corporate debt structure.

(vi) Anti-self-dealing measures, namely, corporate law provisions allowing investors to take legal action against corporate insiders who abuse their position, are systematically linked to debt ownership dispersion. The correlation is not particularly strong in the full sample, but it masks considerable heterogeneity. For all but the firms in the largest size quartile, better anti-self-dealing provisions strongly correlate with higher total debt dispersion, and with higher dispersion in arm's length debt ownership. By contrast, for all firms, and particularly for the larger ones, better anti-self-dealing provisions correlate with lower bank ownership dispersion. That is, the stronger the investor protection, the more dispersed total debt ownership and public bond ownership; but the smaller the number of banks a firm borrows from.

(vii) Anti-self-dealing provisions are systematically linked to corporate debt maturity. This correlation masks considerable heterogeneity, too. Stronger investor protection is correlated with longer maturity of arm's length debt. The effect is largest for

[1] Using the anti-self-dealing index to assess the dispersion of arm's length creditors is exactly in line with the predictions of the theoretical model of Gennaioli and Rossi (2013), who write "The notion of investor protection relevant for this prediction is broader than that of creditor rights and includes shareholder protection against managerial self-dealing and tunneling" (p.624).

the smallest firms, declines monotonically with firm size, and remains strongly significant for all but the largest firm size quartile. Total debt maturity increases with investor protection, but the effect is significant only for small-to-medium size firms. On the other hand, the maturity of bank debt is, if anything, slightly decreasing with investor protection, particularly for larger firms, although the economic significance of this effect is small.

(viii) Creditor rights correlate with the extent to which corporate debt is secured by collateral. The stronger the creditor rights, the larger the likelihood that total debt is secured by collateral. This result is driven entirely by bank debt.

In a nutshell, our results highlight that in countries with stronger investor protection the debt structure features dispersed debt ownership, with dispersed arm's length borrowing coexisting with concentrated bank debt. In countries with weaker investor protection, firms mitigate the unfavorable institutional environment by borrowing at shorter maturity, and by borrowing in USD denominated (foreign) debt. Furthermore, our analysis highlights that investor protection is particularly relevant, indeed crucial, for small-to-medium size firms; whereas the largest firms are able to access international investors–and the protection of US courts–almost irrespective of their country of incorporation. The considerable heterogeneity that we uncover along debt types and firm size thus supports the importance of considering at the same time different sources of debt, including bank loans, credit lines, and arm's length debt, to obtain a full picture of firm's debt structure around the world.

Our results are consistent with legal origin and contracting theories of corporate finance (La Porta et al. (1998), Diamond (2004), Djankov et al. (2007), Djankov et al. (2008a), Gennaioli and Rossi (2013)). Consistent with La Porta et al. (1998) and La Porta et al. (2008), we find that debt ownership dispersion is largest in countries with better investor protection. In particular, our results are consistent with Gennaioli and Rossi

(2013), who rationalize the coexistence of concentrated bank debt and dispersed arm's length debt in the firm's debt structure, and predict that dispersion of total debt and arm's length debt should increase with investor protection.

Our results are also consistent with theories of financial contracting. In particular, our results are consistent with the prediction of Diamond (2004) that firms in poor investor protection countries should mitigate the unfavorable institutional environment by borrowing more short term. Our results are also consistent with the prediction from prior literature that firms in poor investor protection countries should borrow from more banks, e.g., see Diamond (2004) and Bolton and Scharfstein (1996)), although our empirical results highlight that in countries with stronger investor protection arm's length borrowing is more prevalent–and much more dispersed–than bank borrowing.

Related literature. Our paper contributes to the literature on the corporate debt structure. Ongena and Smith (2000) examine bank debt in a cross section of countries; Detragiache et al. (2000) examine the number of bank relationships of Italian firms operating in regions with different enforcement costs; whereas Esty and Megginson (2003), Qian and Strahan (2007), and Bae and Goyal (2009) focus on syndicated loans. These papers find that firms borrow from more banks in countries with weaker investor protection, and interpret their evidence in line with the models of Bolton and Scharfstein (1996) and Diamond (2004) in which firms borrow from multiple creditors to prevent strategic default, and in which strategic default is a more relevant concern in countries with weaker investor protection. We confirm the result from this literature that, *when looking at bank debt in isolation*, bank debt dispersion and the number of banks is larger in civil law countries, and in countries with weaker investor protection; but we show that the picture changes dramatically and we obtain diametrically opposite results when considering total debt ownership, including the ownership of arm's length debt, in line with the predictions of La Porta et al. (1998), La Porta et al. (2008), and Gennaioli and Rossi (2013).

Regarding debt maturity, Demirgüç-Kunt and Maksimovic (1999), Giannetti (2003), and Fan et al. (2012) examine corporate borrowing and find that the fraction of short term (as opposed to long term) debt is larger in countries with weaker investor protection (see John et al. (2017) for an opposing view). More recently, examining syndicated loan contracts Qian and Strahan (2007) and Bae and Goyal (2009) document a more nuanced picture. Qian and Strahan (2007) document that English common law countries have a 1-month longer maturity of bank loans than French civil law firms (see their Table 1), but a slightly shorter maturity in German and Scandinavian legal origin countries; in regression analysis, Qian and Strahan (2007) conclude that firms borrow with longer maturities in countries with stronger creditor protection. Bae and Goyal (2009) reach similar conclusions, although arguing that it is enforceability of contracts that drives longer debt maturity rather than creditor rights. Our analysis builds on Qian and Strahan (2007) who emphasize heterogeneity with respect to firm size; we confirm the importance of heterogeneity with respect to firm size, and in addition we underscore the key role of heterogeneity with respect to debt types. That is, when examining bank borrowing in isolation, we confirm the picture that French civil law countries have a shorter maturity than English civil law countries, although the results are statistically insignificant and their economic magnitude is small. However, when examining arm's length borrowing and total borrowing, the picture changes dramatically, in that firms in common law countries borrow with a much longer maturity, and the result is particularly strong for arm's length debt and for small-to-medium size firms.

Regarding debt classes, John et al. (2021) find that firms from countries with stronger creditor protection choose debt structures with fewer debt types and interpret their result as being in line with predictions of Bolton and Scharfstein (1996) that debt structures should be more dispersed in countries in which strategic default concerns are higher. We show instead that when one examines direct measures of debt ultimate ownership rather than aggregate measures of debt classes, the overall conclusions change dramatically.

Finally, our results contribute to the literature on debt currency, which emphasizes the role of the US dollar as the key anchor of exchange rate regimes, e.g., see Maggiori et al. (2024) and references therein. Our results underscore the importance of borrowing in US dollars for firms located in French civil countries to be able to access the protection of US courts, in line with the legal ruling studied in Hébert and Schreger (2017).[2]

The rest of the paper is organized as follows. Section 2.2 details the data sources and the construction of our final sample. Section 2.3 presents the main patterns of corporate debt structure around the world. Section 2.4 reports our empirical results. Section 2.5 concludes.

[2] A recent and growing empirical literature has examined the effects of financial reform, in particular changes to creditor protection under corporate bankruptcy laws in individual countries, on various firm-level outcomes, including firms' debt structures. See, among others, Scott and Smith (1986), Hackbarth et al. (2015), and Kornejew (2024) for the US, Lilienfeld-Toal et al. (2012) and Vig (2013) for India, Araujo et al. (2012), Assunção et al. (2014), and Ponticelli and Alencar (2016) for Brazil, Cerqueiro et al. (2016) for Sweden, and Rodano et al. (2016) for Italy. While these individual-country studies have been advancing causality by exploiting variation among treated vs untreated firms, they have examined a wide array of heterogeneous reforms and firm-level outcomes; as a result, external validity and comparability of results across countries and reforms may be problematic. We complement these approaches by providing a simple, unified framework across a large number of countries to provide a panoramic analysis of the association between corporate debt structure and legal protection of investors.

2.2 Data and Methodology

We construct a novel dataset on the corporate debt structure of firms around the globe. To that end, we combine data from various data providers. That is, in a first step, we obtain detailed information on the debt structure of international firms from Capital IQ. We then merge this data with quarterly balance sheet information from Worldscope, Compustat Global, and Amadeus. Further, we complement the sample with information on the nature and ownership of bond issues and commercial papers from eMAXX and details on loans and lines of credit, along with information on the lenders of these facilities, from LPC Dealscan. The resulting dataset contains information on the debt structure of 10,136 firms operating across 52 countries from 2002Q2 to 2021Q3. In the following, we discuss in detail the various steps of our sample construction. In what follows we describe in detail the various steps to build our dataset in subsections 2.2.1, 2.2.2, 2.2.3, and 2.2.4, and we discuss our measures of debt ownership concentration in subsection 2.2.5.

2.2.1 Debt Structure Information

As outlined above, we first obtain detailed debt structure information for the universe of firms in *Capital IQ*. As reported in Table 2.1, Capital IQ contains information on 135,807 firms (42.7% of which are public). This data includes the quarterly breakdown of total debt into seven distinct categories: commercial paper, drawn credit lines, term loans, bonds and notes, capital leases, trusts, and a residual category encompassing other types of debt. In our analysis, we rely on the unique company identifier "companyid" provided by Capital IQ which helps us identify duplicate balance sheet information that exist in at least two of Amadeus, Worldscope and Compustat.

2.2.2 Balance Sheet Information

In a second step, we retrieve quarterly balance sheet information of all international firms available in either *Worldscope*, *Compustat Global*, or *Amadeus*. Subsequently, we exclude firm-quarters from our sample which have missing or non-positive values for either total assets or total debt, or reported book leverage outside the unit interval. Moreover, since our ownership data is organized in calendar quarters, we map balance sheet information from fiscal quarters to calendar quarters based on the month in which the fiscal quarter concludes. For example, if a company's fiscal quarter concludes in January, February, or March of 2010, we assign that firm's quarterly balance sheet information to the first quarter of 2010, i.e. 2010Q1.

The merging of Capital IQ with the combined data from Worldscope, Compustat Global, and Amadeus is carried out in several steps. First, we merge the entire Capital IQ dataset with Worldscope. Each firm in Worldscope is matched with its corresponding companyid, as detailed in Appendix Section C. Following the approach outlined by Colla et al. (2020), we then eliminate firm-quarters for which the absolute percentage difference between total debt reported in Worldscope and the sum of debt types reported in Capital IQ exceeds 10%.

Next, we turn to the remaining Capital IQ observations that did not successfully match with Worldscope and attempt to merge them with Compustat Global. Similarly, we merge these observations with Compustat and discard firm-quarters where the discrepancy between total debt reported in Compustat and the sum of debt types reported in Capital IQ exceeds 10% of Compustat total debt. For unmatched Capital IQ entries of European firms, we employ an analog methodology using the financial statements sourced from Amadeus.[3]

[3]In a first step, our integration with Amadeus relies on firms with non-missing ISINs. In a second step, we manually match firms which results in an additional 572 matched firms. Please see the details in Appendix Section C.

Retaining firm-quarter observations with successful merges either in Worldscope, Compustat Global, and Amadeus and eliminating any duplicates leaves us with approximately 2.4 million firm-quarter observations, representing 85,576 unique firms across 145 countries as can be seen in Table 2.1. Notably, about 61% of this sample comprises public firms, with the remaining 39% being private firms.

2.2.3 Debt Ownership Information

Next, we merge our data on debt ownership information. To that end, we obtain direct ownership information of **arm's length debt** (commercial papers, bonds, and notes) from *Thomson Reuters eMAXX*, which offers extensive coverage of institutional investors' fixed-income holdings at the security level. Importantly, these holdings are not limited to publicly traded securities but also include privately placed debt instruments such as 144A and Regulation S bonds, which have become more common in recent years, e.g., see Ellias and De Fontenay (2025). This database primarily includes asset holdings of insurance companies, mutual funds, ETFs, and pension funds.[4] That is, for each calendar quarter, the dataset contains debt ownership data as well as information about the securities themselves and their issuers. When available, we rely on CUSIP as a primary identifier for a security. If CUSIP is not available, we use CINS as an identifier.[5] The first six digits of CUSIP/CINS do not allow to unambiguously identify the issuer for at least two reasons. First, the issuer identifier might change over time due to corporate events. As a result, the 6-digit CUSIP/CINS is different for outstanding securities and new issues. Second, depending on the offering type and market, the same issuer might be assigned several 6-digit identifiers.[6] To overcome this problem, we use Refinitiv Python API and

[4]With some exceptions, the eMAXX database does not include holdings of hedge funds and banks.

[5]The CUSIP International Numbering System (CINS) is a 9-character alphanumeric identifier denoting international securities and is an extension of the CUSIP system used for U.S. and Canadian securities. CINS employs the same nine characters as CUSIP but also contains a letter of the alphabet in the first position, denoting the issuer's country or geographic region.

[6]For example, Microsoft has US bonds with CUSIPs starting with 594918 and 00507V and Eurobonds with CINS that start with U59340.

get the issuer PermIDs of all securities contained in eMAXX. This allows us to uniquely retrieve the "companyid" of the issuing firm from our Worldscope-Capital IQ merge which renders merging of ownership information from eMAXX with data on firm fundamentals straightforward. A more detailed discussion on this procedure can be found in Appendix Section C.

To compile data on **bank debt** ownership, we obtain data from *LSEG Loan Connector* which is part of *DealScan*. This data contains comprehensive historical information on loan characteristics, contract details, terms, and conditions. We focus on the entire DealScan universe to ensure that we capture loans which had been issued before the start of our sample but have not been fully repaid. Unlike eMAXX, which organizes the data as a quarterly panel, DealScan is an event-based dataset. That is, each unique observation corresponds to a distinct combination of $Lender \times Borrower \times Tranche \times Amendment$. Therefore, we first have to re-organize the information contained in DealScan in a quarterly panel. To do so, we assume that, for each tranche, the composition of lenders and the lenders' share in total lending would remain unchanged over time, unless, 1) there is a new amendment to the loan, in which case, we obtain an updated list of lenders; 2) the face value of the loan has been fully repaid; 3) the loan has reached its maturity date. These assumptions can be partially justified by the fact that only around 20% of syndicated loans are traded in secondary markets, as highlighted in (Brealey et al., 2020, Ch. 24, p. 649). In addition, only about one quarter of syndicated loans are eligible for secondary market trading in the US. In Europe and Asia this fraction is even lower.[7] Therefore, the composition of lenders of syndicated loans remains relatively stable over time.

For each quarterly observation of a loan, we calculate the remaining outstanding notional by taking into account the specified repayment schedule. These repayment sched-

[7] As per information from the Bank of International Settlements. Refer to "The Syndicated Loan Market: Structure, Development and Implications" available at www.bis.org/publ/qtrpdf/r-qt0412g.pdf

ules varied in installment frequency, ranging from weekly to once every three years. For cases where the repayment schedule is missing, we either assume a proportional repayment of all the outstanding loans with missing repayment schedules or a bullet payment at maturity. Distinguishing between proportional repayment and bullet repayment is made possible by comparing the observed outstanding Dealscan loans to the corresponding total outstanding loans from Capital IQ. Specifically, if the observed outstanding Dealscan loans amount to less than the total outstanding loans recorded in Capital IQ for the respective observation period, we assume a bullet repayment at maturity for these loans. Conversely, if the outstanding Dealscan loans exceed those in Capital IQ by $x\%$, it is assumed that $x\%$ of these loans have been fully repaid up to the current quarter.

DealScan does not report the type of lenders individually for each loan and instead reports the “union" of all lender types participating in each Tranche[8]. To circumvent this issue, we employ a sequential approach. That is, first we focus on loans with only one lender. For this subsample, we directly observe a one-to-one mapping between lenders and types. Building on this, in a second step, we use data on loans with two lenders where we know one lender’s type from the first step, which allows us to determine the second lender’s type. This iterative procedure allows us to determine the types of 95% of the lenders in DealScan. The remaining 5% of lenders we label as "Other." We then compute lender-specific ownership. For approximately one third of all the loans in DealScan, we have explicit information on each lender’s ownership. For the remaining loans, we assume equal contributions of all participating lenders.

Finally, we merge bank loan ownership information and the existing data. To do so, we follow a two-step process. First, we use the PermID to companyid linking table that we have previously compiled by combining several linking tables provided by Capital IQ and Refinitiv. For a more detailed discussion on this, we refer to Appendix Section C.

[8]The various lender types in DealScan include commercial banks, investment banks, other types of banks, insurance companies, distressed (vulture) funds, mutual funds, pension funds, hedge funds, trusts, and private equity.

Second, we employ the linking table provided by Chava and Roberts (2008).[9] As reported in Table 2.1, combining firm fundamentals with debt ownership data leaves us with 18,104 firms (about 57% are public firms) domiciled in 108 different countries.

2.2.4 Final Data Set

Our final dataset combines information from Capital IQ; quarterly balanced sheet data from Worldscope, Compustat Global, and Amadeus; arm's length debt ownership data from eMAXX; and bank debt ownership data from DealScan. As our paper aims at studying the corporate debt structure around the world, we would like to focus on firms for which we have ownership data for a significant fraction of their total debt. To that end, we apply two more filters in our sample construction which limits our attention to firms for which we identify 1) the debt instruments of at least 50% the book value of the firms' total debt; and 2) the owners of their debt for at least 30% of the book value of the firms' total debt. That is, in summary, a hypothetical firm A is included in our sample in quarter t if and only if the following criteria are met:

1. Capital IQ contains data on the debt structure of firm A for quarter t;

2. Worldscope, Compustat Global, or Amadeus contain balance sheet data of firm A for quarter t;

3. The relative percentage difference between the total debt reported in Capital IQ and the balance sheet is less than 10%;

4. Debt instruments identified through eMAXX and DealScan account for at least 50% of firm A's total debt amount outstanding according to the balance sheet;

[9] The linking table of Chava and Roberts (2008) is based on the legacy version of DealScan. Therefore, the borrower company identifiers must be updated according to the revamped version.

5. Debt ownership must be known for at least 30% of firm A's total debt amount outstanding according to the balance sheet.

After applying these criteria, we exclude all observations from countries for which we have data on fewer than ten distinct firms. As reported in Table 2.1, our final sample includes 127,179 firm-quarter observations for 10,136 unique (about 55% are public firms) firms in 52 countries.

It is important to note that, despite merging several data sets from different sources and applying many filters when constructing the data, our final sample does not look that different than the sample of firms which we used as "population", i.e., the universe of firms contained in Capital IQ. As reported in Table 2.1, for example, the share of public firms is 55% in our final sample which compares to 43% in the Capital IQ sample. Given the fact that data availability is typically much better for public compared to private firms, this increase in the share of public firms seems quite small. Moreover, the slight increase in the share of arm's length debt and credit lines in the final sample compared to the starting sample suggests that our sample features relatively more large firms compared to the population. Again, this shouldn't come as a surprise as generally, data is more easily available for larger firms. Importantly, however, applying our filters four and five above does not materially change neither the average debt structures nor the average leverage of our sample firms.

Figure 2.1 shows the global average capital structure and debt structure for our pooled sample, including the composition of various debt types. The debt categories comprise arm's-length debt, bank loans, credit lines, and other forms of debt. Figure 2.1 documents average leverage, measured as total book debt divided by total assets, is 32.2%. When examining debt types as a percent of total book debt, bank loans are 34.8%, arm's length debt is 29.2%, credit lines are 21%, and other types collectively amount to 15.1%.

Figure 2.2 reports the global evolution over 2002-2021 of leverage and debt structure. It shows that average leverage is quite stable over 2002-2021, hovering around 30-to-33%

of total assets. On the other hand, in terms of debt structure over time we observe a gradual substitution of bank loans and–to a lesser extent–arm's length debt for credit lines. In fact, credit lines steadily decline from about 30% of total debt in 2002 to about 15% in 2021, whereas bank loans increase from 27% to 36% and arm's length debt from 27% to 32%.

2.2.5 Measures of Ownership Concentration

To quantify the extent of concentration in debt ownership, we rely on four well-known measures of ownership concentration, originally developed in the literature on equity ownership. That is, we use for our analysis the number of creditors, the minimum number of creditors required to cross 10%, 20%, and 30% of debt ownership, the Herfindahl-Hirschman Index (HHI), and the cumulative ownership of the one, three, five, and ten largest lenders.

Our first concentration measure, the *number of total creditors*, is relatively uncontroversial both to measure as well as to interpret (see, for example, Li and Yu (2022)). That is, a lower (higher) number of creditors indicates a higher (lower) degree of ownership concentration.

Second and relatedly, the *minimum number of creditors required to cross a certain ownership threshold* is defined as follows:

$$NtC_{X\%} = min\{n\ s.t. \sum_{i=1}^{n} \omega^i \geq X\%\} \tag{2.1}$$

where NtC stands for "Number to Cross", ω^i denotes the cumulative ownership share in percent of the i-th largest lender, and $X\%$ denotes the ownership threshold. For the ownership threshold we consider 10%, 20%, and 30%, respectively. A low (high) number of creditors to cross a certain ownership threshold indicates higher (lower) degree of ownership concentration (Franks et al., 2009).

Third, the *Herfindahl-Hirschman Index (HHI)* represents the sum of squared fractional holdings of creditors, calculated relative to the total debt for which we observe ownership in a given firm and quarter. That is,

$$HHI = \sum_{i=1}^{n} {\omega^i}^2, \tag{2.2}$$

where n is the total number of observed creditors and ω^i denotes the ownership share of i-th largest lender as a fraction of total debt for which we observe ownership. That is, a high (low) HHI indicates high (low) degree of ownership concentration. This methodology aligns with previous studies that rely on observations where less than 100% of the underlying ownership is observed (see, for example, Li and Yu (2022)). HHI for bank debt and bond debt (also referred to as arm's length debt) is defined in an analogously. That is, the denominator of fractional holdings are total bank debt with observed ownership and total arm's length debt with observed ownership, respectively.

Importantly, using the total amount of underlying debt in the denominator of fractional holdings, rather than the total amount for which we ownership, would inevitably bias the resulting HHI downward. The magnitude of this bias is directly related to the fraction of debt with unobserved ownerships. Instead, we derive an unbiased estimator under the assumption that the ownership concentration we observe is representative for the ownership concentration of all creditors of a firm. The fifth criterion in our final sample construction guarantees that we retain only those observations for which we have full information on a substantial fraction of debt. Moreover, when calculating HHI statistics for bank and bond debt separately, we apply a similar criterion. Specifically, we calculate the HHI for bank and arm's length debt only for firm-quarters for which we have ownership information for at least 30% of total arm's length debt.

Finally, we calculate the cumulative ownership measures in line with Franks et al.

(2009) as follows:

$$C1 = \omega^1 \tag{2.3}$$

$$C3 = \sum_{i=1}^{3} \omega^i \tag{2.4}$$

$$C5 = \sum_{i=1}^{5} \omega^i \tag{2.5}$$

$$C10 = \sum_{i=1}^{10} \omega^i \tag{2.6}$$

where, as above, ω^i denotes the ownership share in percent of the i-th largest lender. That is, a higher (lower) value for the measures $C1$ to $C10$ indicates a higher (lower) ownership concentration. Analogous to the HHI, the ownership shares are relative to the total amount of debt for which we have ownership information. Moreover, for statistics specific to bank or arm's length debt we require to observe ownership information for at least 30% of total bank or arm's length debt outstanding to avoid any downward bias. For a more detailed discussion of our variable definitions, please see Appendix Table A.

2.3 Patterns

We aim to provide the most up-to-date characterization of corporate debt structure around the world. In this Section, we report descriptive statistics of corporate debt structure by legal origin.

2.3.1 Summary Statistics

Figure 2.3 provides a map of corporate debt structure around the world for four legal origins, namely English common law, Scandinavian civil law, German civil law, and French civil law, following the classification in La Porta et al. (2008). The map illustrates the considerable heterogeneity in debt structure around the world. Two clear patterns emerge: 1) English common law firms borrow predominantly using arm's length debt; 2) Civil law firms borrow predominantly using bank loans.

To dig deeper, Table 2.2 reports detailed summary statistics for our final sample, for each country individually and for the four legal origins. Average leverage is quite similar across legal origins, about 33-to-34% of assets in both French civil law and English common law countries, and not far from that in German and Scandinavian countries. Table 2.2 substantiates the considerable heterogeneity in debt structure around the world illustrated in Figure 2.3. In fact, firms in English common law countries borrow predominantly using arm's length debt at 36% of total debt, followed by bank loans at 26% and credit lines at 24%. On the other hand, bank loans amount to 55% of total debt in French civil law countries and even 62.6% in German civil law countries. By contrast, arm's length debt amounts to 16% of total debt in French civil law countries and a meager 3% in German civil law countries.

The last two columns of Table 2.2 show that we identify on average about 80% or more of firms' debt instruments and we have ownership data on average for about 70% of firms' total debt outstanding. This gives us confidence that our data is rich enough to make precise statements about the firms included in our final sample. About 45% of the firms in our sample are in the US (4449 firms). For this reason, we perform extensive robustness tests excluding US firms, and we find very similar results to those we report here using the pooled sample. Finally, while we observe both public and private firms in all countries, the share of public firms is much lower in English common law countries compared to the other three legal origins.

2.3.2 Debt Ownership Concentration

Figure 2.4, Panels A, B, and C, illustrates the differences in the distribution of debt ownership concentration between English common law and French civil law countries using the number of creditors measure, separately for bank debt, arm's length debt, and total debt, respectively.

Panel A shows that firms in French civil law countries borrow from more banks than firms in English common law countries. The interpretation is that bank debt ownership is more concentrated in English common law countries than in French civil law countries. This pattern is in line with prior literature that examine bank lending across countries such as Esty and Megginson (2003), Qian and Strahan (2007), and Bae and Goyal (2009).

Panel B shows that the pattern dramatically reverses when considering ownership of arm's length debt, as firms in English common law countries borrow from many more bondholders than firms in French civil law countries, implying firms in English common law countries have more dispersed arm's length debt ownership than firms in French civil law countries. Panel C shows that this pattern carries over to total debt ownership concentration, as firms in English common law countries borrow from many more overall creditors, and have thus much more dispersed debt ownership, than firms in French civil law countries.

Figure 2.5 and Figure 2.6 reports the exact same patterns using the C5 and the $NtC_{30\%}$ measures of debt ownership concentration. Firms in French civil law countries have more dispersed bank ownership than firms in English common law countries, as the collective debt ownership of the top 5 banks is smaller in French civil law countries than in English common law countries, and the number of banks needed to cross the 30% debt ownership threshold is larger in French civil law countries than in English common law countries, consistent with prior literature. However, this pattern dramatically reverses when examining arm's length debt ownership and total debt ownership: firms in English common law countries have a much more dispersed ownership of arm's length debt and of total debt than firms in French civil law countries, also using the C5 and $NtC_{30\%}$ measures.

Figure 2.7 examines the evolution over the firm's life cycle of leverage (Panel A), debt structure (Panel B), and debt ownership concentration (Panel C). Here, year 0 denotes the year the firm is established. The Figure shows that while leverage is quite stable over the firm's life cycle (Panel A), debt structure displays a substitution of arms' length debt for bank loans and credit lines, as arm's length debt starts picking up at about year 60 of the firms' life (Panel B). Panel C shows that as soon as arm's length debt starts picking up, the firms' debt ownership concentration also decreases sharply, at least in terms of overall number of creditors, and number of creditors needed to cross the 30% threshold.

Table 2.3 reports bank debt ownership concentration statistics by individual countries and legal origins. It shows that both the number of loan tranches as well as the number of loans is smaller for firms in English common law compared to firms in civil law legal origins. The number of lead arrangers in syndicated loans is also lower in firms in English common law countries (on average about 3) compared to firms in French civil law countries (on average about 6.3).

Table 2.3 also reports a larger set of measures of bank debt ownership concentration, and they confirm the patterns shown above for number of creditors, C5, and $NtC_{30\%}$ also for other measures. For example, we find for firms in English common law countries compared to firms in French civil law countries a higher HHI (25.05% vs 21.47%) and lower numbers required to cross the 10% (1.18 vs 1.19), 20% (1.76 vs 1.81), and 30% (2.43 vs 2.55) ownership thresholds. That is, all measures point to the fact that the ownership concentration of bank debt is on average *higher* for English common law countries, consistent with prior literature.

Our main innovation is to examine together debt ownership not only of bank loans and credit lines but also of bonds and in general of arm's length debt. We show that by doing so the picture changes dramatically relative to prior literature, reversing the conclusions reached so far. Table 2.4 shows that the patterns on the ownership concentration of arm's length debt are diametrically opposite to the patterns for bank debt. That is, the number of overall bondholders and the minimum number of bondholders required to pass a cumulative ownership of 10%, 20%, or 30%, respectively, are all clearly higher for English common law countries. The economic magnitude is stark. For example, the average number of bondholders across firms in English common law countries equals 163 and compares with the average of about 23 bondholders across firms in civil law countries. Consistent with this clear empirical pattern, the average HHI for firms in English common law countries is well below the average HHI for firms in French civil law countries (5.05% vs 8.16%). Therefore, we find that bond debt ownership in firms in English common law countries is much *less* concentrated than in firms in civil law countries.

This finding is the exact opposite of what we just documented for bank loan ownership concentration above. Notably, the English common law statistics are quite stable throughout a number of countries, including Canada, Bermuda, Ireland, Jersey, United Kingdom, and the United States. In these countries, on average, a) bond debt makes up for an economically meaningful fraction of overall debt (26% for Ireland to 42% for the United States), and b) we observe ownership information for a significant fraction of total bond debt (42% for Jersey to 61% for the United States). At the same time, the results are not driven by US firms, as we obtain very similar results if we exclude US firms from the sample. These exercises alleviate the concern that the common law results are driven entirely by a single country or a few outliers.

Next, we show how bank and arm's length debt aggregate to total debt and, thus, we document ownership concentration of total debt of firms. Table 2.5 reports statistics by individual countries and by legal origins. Importantly, across all our sample countries, on average, we have debt ownership information for about 73% of total debt (this percentage is lowest for Colombia with 52.60% and highest for New Zealand with 87.47%). This observation suggests that our empirical evidence is representative for the corporate debt concentration of our sample firms. Overall, we find that debt ownership in firms in English common law countries is on average *less* concentrated than in firms in other legal origins. This conclusion is consistently backed up by all of our four measures of ownership concentration when comparing average values across English common law and French civil law countries. In English common law countries the 1) the number of all creditors is higher (173 vs 66), 2) the minimum number of creditors to cross the ownership thresholds of 10%, 20%, and 30% are higher (1.51 vs 1.27, 2.75 vs 2.04, and 4.42 vs 3.02), 3) the HHI is lower (17.79% vs 18.90%), and 4) the cumulative ownership of the 1, 3, 5, and 10 largest creditors are lower (22.79 vs 23.82, 36.16 vs 40.11, 52.09 vs 60.03, and 69.66 vs 80.84).[10]

To sum up, our data confirms the results about ownership dispersion of bank debt from the literature. Importantly, however, we find that the patterns of ownership dispersion of total debt are diametrically opposite. When considering arm's length debt and thus total debt, our empirical evidence points to the striking fact that ownership dispersion of total debt and arm's length debt is much more dispersed in firms in English common law countries than in firms in other legal origins, that is, exactly the opposite of what was documented regarding bank debt. Intuitively, total debt ownership dispersion is driven by the dispersion of arm's length debt because of the quantitative importance of arm's length debt in the debt structure of our sample firms, and particularly of firms in English common

[10] The results are even stronger when comparing firms in English common law countries with firms in German civil law countries.

law countries. Therefore, our comprehensive data allows us to paint a much more nuanced picture of debt ownership dispersion compared to existing studies which exclusively relied on bank debt to make inference about debt ownership dispersion, including Ongena and Smith (2000), Esty and Megginson (2003), Qian and Strahan (2007), and Bae and Goyal (2009).

Debt Maturity

Figure 2.8 reports the distribution of debt maturity, separately for bank debt (Panel A), arm's length debt (Panel B), and total debt (Panel C), comparing firms in French civil law countries and firms English common law countries. For each firm, we obtain a single measure of maturity based on weighted average of the maturity of the relevant debt instruments outstanding. Panel A shows that bank debt maturity is slightly shorter in firms in French civil law countries relative to firms in English common law countries, although the magnitude of the difference is small. Panel B shows that arm's length debt maturity is much shorter in firms in French civil law countries relative to firms in English common law countries, and Panel C confirms this pattern for total debt maturity.

Table 2.6 reports detailed statistics for debt maturity both for individual countries and for legal origins. The average maturity across all sample countries is 4.88 years for total debt, 7.78 years for arm's length debt, and 3.05 years for bank debt. Importantly, however, there is substantial cross-sectional variation in debt maturity across countries and, in particular, across legal origins. For example, total debt maturity is substantially longer (5.39 years) for firms in English common law countries compared to firms in civil law countries (3.40 years).[11] Consistent with this, the maturity of bond debt is also longer in firms in English common law countries compared to firms in civil law countries (8.19 years vs 4.34 years). Interestingly, however, the pattern is reverted for bank debt.

[11]As before, we calculate the average of statistics across civil law origins by weighting origin-specific averages by number of firms in a given origin. For example, the average total debt maturity of 3.40 years is calculated as follows: $(319 \times 3.40 + 1920 \times 3.20 + 1077 \times 3.77) / (3316)$.

That is, firms in English common law countries have a slightly shorter bank debt maturity compared to civil law countries (3.00 years vs 3.17 years), although the magnitude of such difference is small and not statistically significant. In the context of our findings from Section 2.3.2, we can therefore summarize that debt ownership concentration is negatively correlated with debt maturity. Our data indicates this relationship is robust no matter whether we focus on bank, arm's length, or total debt. This correlation supports the idea in Diamond (2004) that debt maturity and debt ownership dispersion are substitutes.

2.3.3 Debt Collateral Security

Figure 2.9 reports the distribution of the percentage of bank debt (Panel A), arm's length debt (Panel B), and total debt (Panel C) that is secured by some collateral, comparing firms in French civil law countries and firms in English common law countries. The Figure shows that a larger fraction of debt is secured by some collateral for firms in English common law countries relative to firms in French civil law countries, and this is true for both bank debt and arm's length debt.

Table 2.7 reports detailed statistics on the fraction of secured debt in individual countries and legal origins. Overall, about one third (32.31%) of total debt is secured by some form of collateral. Notably, the fraction of secured debt is much higher for English common law countries compared to civil law countries (35.16% vs 24.18%). That said, however, the fraction of secured debt for firms from Scandinavian legal origin is as high as the one across firms in English common law countries. Moreover, we examine the fraction of secured debt that is due to bank and arm's length debt, respectively. Perhaps not surprisingly, secured debt is mostly composed of secured bank debt as opposed to secured arm's length debt. This is true across all our sample countries. More interestingly, the composition of secured debt into bank and bond debt exhibits interesting cross-sectional variation. In fact, we find that the share of bond debt is significantly higher in firms in English common law countries compared to firms in civil law countries (11.08% vs 3.57%).

The flip side of this statement implies that the share of bank debt is lower in firms in English common law countries compared to firms in civil law countries (88.92% vs 96.43%). To sum up, we find that debt ownership concentration is negatively correlated with the fraction of secured debt.

2.3.4 Debt Currency

Figure 2.10 displays the distribution of debt currency, comparing firms in French civil law countries with firms in English common law countries. The Figure shows that firms in French civil law countries are much less likely to borrow in local currency and much more likely to borrow in USD-denominated debt as a foreign currency than firms in English common law countries.

Table 2.8 reports detailed statistics for firms' debt currency choices in individual countries and all legal origins. Overall, we find that firms operating in common law countries are more likely to issue debt in their local currency compared to firms in civil law countries. That is, while 89.93% of firm's debt in English common law countries is issued in local currency, this is the case for only 77.77% of firm's debt in civil law countries. On the other hand, firms in French civil law countries borrow 37.87% in USD denominated arm's length debt and 33.98% in USD denominated bank debt. These findings are consistent with the idea in Mellor et al. (2024) and Thiel (2000) that in international transactions contracting parties may attempt to mitigate an unfavorable institutional environment by borrowing in US dollars, which often comes with the protection of New York law and US courts.

2.4 Results

2.4.1 Legal Origin, Investor Protection, and Debt Structure

We examine the empirical patterns documented from Section 2.3 in a multivariate context at the firm level. That is, we run the following regression models

$$y_{i,j,c,t} = \beta_1 \times \mathbb{1}\text{French}_c + \beta_2 \times \mathbb{1}\text{German}_c + \beta_3 \times \mathbb{1}\text{Scandinavian}_c + \delta' \mathbf{X}_{i,j,c,t} + \eta_j + \eta_t + \varepsilon_{i,j,c,t}, \tag{2.7}$$

where i indexes firms, j industries, c countries, and t year-quarter, the independent variables y are measures of ownership concentration, debt maturity, secured debt, or debt currency. Further, $\mathbb{1}$French, $\mathbb{1}$German, and $\mathbb{1}$Scandinavian are dummies which equal one if a firm belongs to the French, German, or Scandinavian legal origin, respectively. In addition, $\mathbf{X}$ denotes a vector of control variables including firms' assets and age, and GDP per capita. Finally, η_j and η_t denote industry and year-quarter fixed effects.

To explore how our results relate to country-specific measures of legal protection and, thus, exploit more granular variation, we rely on the anti-self-dealing index from Djankov et al. (2008b), *ASD*. The anti-self-dealing index measures the legal protection against expropriation by corporate insiders. The index is available for most of our sample countries and focuses on private enforcement mechanisms, such as disclosure, approval, and litigation, that govern a specific self-dealing transaction. The index exhibits substantial cross-sectional variation across our sample countries as can be seen from Appendix Table A2. On average, the index is higher for common law countries compared to civil law countries. That is, we also run the following regression models:

$$y_{i,j,c,t} = \beta \times ASD_c + \delta' \mathbf{X}_{i,j,c,t} + \eta_j + \eta_t + \varepsilon_{i,j,c,t}, \tag{2.8}$$

and

$$y_{i,j,c,t} = \beta \times ASD_c \times \text{size group}_{i,j,c,t} + \delta'\mathbf{X}_{i,j,c,t} + \eta_j + \eta_t + \varepsilon_{i,j,c,t}, \tag{2.9}$$

where *size group* denotes four groups formed based on the quartiles of firm size within a country, with firm size measured as total assets in US Dollars. The size quartiles serve two important purposes: First, they categorize firms into size groups within the population, rather than in our final sample[12]. This ensures that even if the final sample is skewed towards larger firms due to data availability, we can still mitigate concerns about a potential bias by studying different segments of the true spectrum of firms in the economy.

Second, the interaction between legal protection and firm size allows us to explore the heterogeneous effects of investor protection across different size groups. This choice is further motivated by the observation that many of the largest firms are international, and for these firms, it is often possible to access the protection of New York law and of US courts, almost irrespective of where they are located.

Debt Ownership Concentration

Table 2.9 reports the regression results when we use the number of creditors, bondholders, banks, and lead arrangers as independent variables. Even when controlling for size, age, GDP per capita and fixed effects, we find that total debt ownership in civil law countries is more concentrated compared to English common law countries. That is, the point estimates for all three dummy variables are negative and, for French and Scandinavian legal origins, highly statistically significant. The same findings are true when we exclusively focus on arm's length debt. That is, the number of total creditors and arm's length creditors are significantly lower for civil law countries. In contrast, in the case of bank debt we find that both the number of banks as well as the number of lead arrangers are

[12]The first time in our data preparation when we have firm size is after we merge the data in Capital IQ with our balance sheet information in Compustat, Worldscope, and Amadeus. This is the point where we form the size quartiles.

higher in civil law countries. That is, bank debt ownership is less concentrated in civil law countries. Overall, our regression results are entirely consistent with the unconditional sample averages presented in Section 2.3.2. Therefore, our results are unlikely driven by individual outliers in our sample.

Importantly, our results also hold when we allow for more granular variation in the variable that relates to legal system of our sample countries. In fact, the interpretation of the results is qualitatively the same when we employ the anti-self-dealing index from Djankov et al. (2008b) and run the regression models from equations (2.8) and (2.9). That is, the regression coefficient of the anti-self-dealing index is positive for the number of total debt and arm's length debt creditors but negative for the number of banks and lead arrangers. The lack of statistical significance in these coefficients hints to the possibility that the relationship between legal environment and debt ownership concentration might not be equally strong for all firms within a certain country. For example, one would expect country-specific legal protection for creditors to matter more for smaller, locally active firms and less so for large, international firms which have access to international credit markets. Columns (9) and (10) of Table 2.9 confirm this conjecture in the data. That is, we find a highly statistically significant and positive relationship between the anti-self-dealing index and the number of total debt or bond creditors for all firms except the very large firms, i.e., firms in the highest quartile of the country-specific size distribution in the population. In terms of economic significance, a one standard deviation increase in the anti-self-dealing index corresponds to an increase of 118.5, 57.6, 18.9 in the number of creditors, respectively for small, medium, and large firms. This effect is economically large given the unconditional sample average of about 136 creditors. The effect on very large firms is negligible. Moreover, we see the monotonically decreasing point estimates of the ASD-coefficient from the smallest to largest size group as further confirmation of our prior. Notably, the pattern changes in columns (11) and (12) of Table 2.9 for the number of banks and lead arrangers.

In Table 2.10 we repeat the analysis from Table 2.9 for an alternative measure of debt ownership concentration, namely the number of creditors required to cross an ownership threshold of 30%.[13] All our findings are robust to this alternative specification. Overall, debt ownership concentration is positively correlated with legal protection. Further, this positive relationship is strongest for small firms and gets gradually weaker with firm size. That is, a one standard deviation increase in the anti-self-dealing index corresponds to an increase of about 0.5 to 1 in the number of creditors required to cross 30% of ownership. This is an economically significant increase given that the unconditional average is about 4. For further robustness, Appendix Tables A3 and A4 tabulate regression results for the two remaining measures of debt ownership concentration, namely the cumulative ownership measure, $C10$, and HHI. Again, all our results are robust to these alternative specifications.

Discussion of Debt Ownership Concentration Results

Our results accord well with the empirical predictions of the theoretical model of Gennaioli and Rossi (2013), who predict that the "dispersion of noncontrolling creditors should be observed at relatively high levels of investor protection" (p. 604). This prediction is the opposite of the traditional idea that debt dispersion makes lenders tough following a strategic default (Bolton and Scharfstein (1996)), implying that firms should borrow from more lenders when enforcement cost is higher (Diamond (2004)). It is useful to discuss briefly here the sources of these different theoretical predictions (the reader can of course always read the original contributions).

Gennaioli and Rossi (2013) study theoretically the optimal debt structure to resolve financial distress, and outline under which conditions such debt structure yields an efficient resolution of financial distress. Their model is in the tradition of Aghion and Bolton (1992) and Hart and Moore (1998), in that debt contracts can allocate control rights

[13] Note that all results remain qualitatively similar when we use the number of creditors required to cross an ownership threshold of 10% or 20%, respectively.

over the reorganization versus liquidation decision, and cash flows are less pledgeable than liquidation proceeds. Unlike Aghion and Bolton (1992) and Hart and Moore (1998), Gennaioli and Rossi (2013) allow for multiple creditors.

Gennaioli and Rossi (2013) find that the optimal debt structure consists of two classes of creditors. One class is concentrated on a large creditor (a 'bank') who has exclusive control over the liquidation versus reorganization decision. This creditor is given both physical collateral and an equity stake in the reorganized firm. Crucially, this equity claim is undercollateralized, in the sense that it pledges a share of physical assets that is smaller than the creditor's equity stake. The other debt class is dispersed among small, atomistic creditors who have no control rights. These creditors are entitled to obtain some liquidation proceeds but are 'wiped out' in reorganization.

The logic of this debt structure is insightful. Giving the large creditor equity in the reorganized firm removes the creditor's liquidation bias. In addition, under-collateralization makes him fully internalize the upside of reorganization. Under-collateralization, the novel feature of Gennaioli and Rossi (2013), works best in the presence of a second debt class: distributing the remaining liquidation proceeds to a debt class without control rights, as opposed to 'leaving them on the table,' maximizes total repayment to all creditors. The debt class of non-controlling creditors should then be dispersed to prevent collusion, which would reintroduce a liquidation bias. The presence of two debt classes differing in their control rights rationalizes the coexistence of 'bank' and 'arm's-length' debt in a firm's debt structure. As a result, Gennaioli and Rossi (2013) predict that debt ownership dispersion should increase with investor protection, and in particular "The notion of investor protection relevant for this prediction is broader than that of creditor rights and includes shareholder protection against managerial self-dealing and tunneling" (Gennaioli and Rossi (2013), p.624).

This notion differs from the traditional idea that debt dispersion makes lenders tough following a strategic default, e.g., Bolton and Scharfstein (1996), and Diamond (2004), and as a result one should expect firms to borrow from more creditors when the institutional environment is weaker. The key driver of the different predictions is that, unlike Gennaioli and Rossi (2013), neither Bolton and Scharfstein (1996) nor Diamond (2004) study the allocation of control rights over the liquidation decision.

Debt Security

Next, we examine how creditor protection correlates with the firms' choice to issue debt secured by collateral rather than unsecured debt. To do so, we rely on the creditor rights index from Djankov et al. (2007) rather than the anti-self-dealing index used before in regression specifications otherwise similar to equations 2.8 and 2.9. In fact, the creditor rights index measures directly the rights of secured creditors in bankruptcy, and it is therefore more appropriate to examine cross-country variation in secured debt.

Table 2.11 reports regression results when we use a firm's fraction of secured debt as an independent variable. Two main findings emerge. First, firms operating in civil law countries are significantly less likely to issue secured debt compared to common law country firms. This finding is consistently true for total debt, arm's length debt, and bank debt. Perhaps not surprisingly, the results are weakest for bond debt which rarely happens to be secured by collateral, as can be seen from Table 2.7. Consistent with these findings, we estimate a highly statistically significant positive association between the creditor rights index and the tendency to issue secured debt.

Second, our results show that the strong and positive association is entirely due to the large and very large firms in our sample. In fact, the association between creditor rights and issuance of secured debt is even negative for firms in the smallest size group. That is, while creditor rights and secured debt seem to be substitutes for small firms, they appear to be complements for large firms. Our estimates are economically meaningful. For example, for total debt, a one standard deviation increase in the creditor rights index corresponds to a 4.98 percentage point increase in the fraction of secured debt. At the unconditional sample average, this implies an increase of about 16% in the fraction of secured debt.

Debt Maturity and Currency

Table 2.12 reports the regression results for debt maturity. Overall, the reported results suggest that better legal protection results in longer maturity for arm's length debt. That is, firms in civil law countries tend to issue debt with shorter maturity compared to firms in common law countries. As usual, bank debt behaves differently to bond debt. Therefore, bank debt maturity is longer in countries that have lower legal protection as measured by the anti-self-dealing index from Djankov et al. (2008b). In terms of economic significance, a one standard deviation increase in the anti-self-dealing index corresponds to an increase of 1.06, 0.64, 0.42, and 0.36 years in the maturity of arm's length debt, respectively for small, medium, large, and very large firms. Yet, these effects are partly offset by the simultaneous decrease in the maturity of bank debt, such that the maturity of total debt is only significantly positively correlated with legal protection for small-to-medium-size firms.

In Table 2.13 we explore the relationship between legal creditor protection and firms' debt currency choice. Overall, the regression results suggest that civil compared to common law countries issue to lesser (higher) extent debt denominated in the local currency (in US Dollars). Moreover, the signs of our coefficient estimates of the anti-self-dealing index indicate that small firms operating in low creditor protection countries are more likely to issue debt denominated in US Dollars rather than in their local currency. This pattern suggests that issuing US Dollar denominated debt is a (imperfect) substitute for legal creditor protection. While our data is too noisy to recover precise statistically significant estimates, the sign and relative size of the coefficient estimates are in line with our interpretation of the data.

Overall, our results indicate that two ways by which firms try to mitigate a weak legal protection is by issuing debt with a short time to maturity and denominated in US Dollars. These results are in line with the prediction by Diamond (2004) that debt maturity should negatively correlate with enforcement costs, and with the general idea that USD denominated debt may come with the protection of New York law and of US courts, e.g., see Hébert and Schreger (2017) and Mellor et al. (2024). Therefore, our results point to a trade-off between debt ownership concentration on the one hand, and debt maturity and US dollar denomination of debt contracts on the other hand.

2.5 Conclusion

Combining different datasets for the first time and conducting manual checks for thousands of firms, we construct a new data set on corporate debt structures around the world. Our database pieces together the debt structure of 10,136 public and private firms from 52 countries between 2002 and 2021, comprising 124,101 corporate bond issues held by 63,323 bondholders and 85,817 bank loans and credit lines extended by 10,154 banks.

There are three main parts of our analysis. First, we provide an anatomy of corporate debt structure around the world. Corporate leverage is remarkably similar across the world at about one third of the firm's assets. Bank loans are the most common form of debt financing, followed by arm's length debt and credit lines, but there is a lot of heterogeneity.

Second, we examine the association between debt structure (and debt ownership concentration) and legal origin. Total debt ownership is most concentrated in civil law countries, and most dispersed in common law countries. This is a novel result in the literature. There is considerable heterogeneity: In common law countries the debt structure features dispersed debt ownership, with dispersed arm's length borrowing coexisting with concentrated and secured bank borrowing. In civil law countries firms mitigate the unfavorable institutional environment by borrowing from more banks (but far fewer arm's length creditors), at shorter maturity, and by borrowing in USD-denominated (foreign) debt. These patterns are strongest for small-to-medium size firms, whereas the largest firms are able to achieve dispersed debt ownership by international investors (and the protection of US courts) almost irrespective of their country of incorporation.

Third, we associate corporate debt structure with institutional characteristics related to legal origin. We find that provisions that protect small outside investors from the self-dealing activities of managers or dominant insiders are significant correlates of corporate debt structure. Stronger protection against self-dealing correlates with higher total debt dispersion, and with higher dispersion in arm's length debt ownership. This pattern holds for all but the largest firms. By contrast, for all firms, and particularly for the larger ones, better anti-self-dealing provisions correlate with lower bank ownership dispersion. Stronger protection against self-dealing also correlates with longer maturity of arm's length debt. The effect is largest for the smallest firms, declines monotonically with firm size, and remains strongly significant for all but the largest firms. We also find that rights of secured creditors in bankruptcy are significant correlates of the likelihood of debt being secured by collateral. The stronger the creditor rights, the larger the likelihood that total debt is secured by collateral. This result is entirely driven by bank debt.

These results support the key insights of the law and finance literature. Consistent with La Porta et al. (1998) and La Porta et al. (2008), we find that debt ownership dispersion is largest in countries with better investor protection. In particular, our results are consistent with Gennaioli and Rossi (2013), who rationalize the coexistence of concentrated bank debt and dispersed arm's length debt in the firm's debt structure, and predict that dispersion of total debt and arm's length debt should increase with investor protection. Our results on debt maturity are consistent with the prediction of Diamond (2004) that firms in poor investor protection countries should mitigate the unfavorable institutional environment by borrowing more short term. Overall, our results highlight the substitutability of debt ownership concentration with debt maturity and currency. In better institutional environments, firms are more likely to borrow from many dispersed creditors at longer maturities and in local currencies, whereas in worse institutional environments firms are more likely to borrow from fewer creditors at shorter maturities and in USD denominated debt contracts.

We view this paper as a first step in reassessing fundamental questions in corporate finance, related to the determinants and consequences of corporate debt structure. Our analysis abstracts from the exact role of banks and bondholders on how to resolve financial distress, on how to renegotiate or restructure debt, on whether to liquidate corporate assets and how to allocate the proceeds, in or outside of formal bankruptcy proceedings. Banks and bondholders might be relatively passive or play a more active role, on their own or in coalition with others. Future work should examine in detail the mechanics of resolving financial distress when the corporate debt structure comprises many heterogeneous creditors in a number of separate debt classes, and should examine in detail the international dimension of corporate debt structure and the resolution of financial distress.

2.6 Tables

Table 2.1: Sample Construction

Sample	Legal Origin	Num Countries	Num Firms	Num Firms %	Num Public Firms	Num Zero Leverage Firms	Num Firm Quarters	Leverage	Arm's Length Debt %	Loans %	Lines %	Other Debt Types %
CIQ	All	184	135807	100.0	57940	751	3563609		21.5	46.6	13.0	19.1
CIQ	English Common Law	39	75655	55.7	32525	623	1798396		26.6	35.5	15.5	22.4
CIQ	Scandinavian Civil Law	4	3731	2.7	1591	11	106464		25.2	44.6	8.9	21.4
CIQ	German Civil Law	20	38182	28.1	17126	82	1105415		13.2	61.7	10.1	15.1
CIQ	French Civil Law	58	16965	12.5	6180	26	518301		22.5	49.0	12.2	16.5
CIQ	Other	63	1274	0.9	518	9	35033		16.1	52.4	12.4	19.3
CIQ & BS	All	145	85576	100.0	52063	254	2422026	0.25	17.1	50.5	15.2	17.6
CIQ & BS	English Common Law	35	47920	56.0	27516	192	1244098	0.26	22.9	37.8	18.2	21.2
CIQ & BS	Scandinavian Civil Law	4	2678	3.1	1563	6	78040	0.25	16.6	52.7	10.3	21.0
CIQ & BS	German Civil Law	20	23150	27.1	16541	46	729402	0.23	10.4	65.0	11.4	13.5
CIQ & BS	French Civil Law	54	11159	13.0	6012	9	349096	0.27	14.7	56.6	15.1	14.5
CIQ & BS	Other	38	739	0.9	469	1	21390	0.25	13.5	52.9	14.3	19.5
CIQ & BS & Ownerships	All	108	18104	100.0	10379	0	388839	0.35	37.9	38.5	14.0	9.7
CIQ & BS & Ownerships	English Common Law	34	10340	57.1	4954	0	215968	0.35	45.2	27.2	18.2	9.5
CIQ & BS & Ownerships	Scandinavian Civil Law	4	630	3.5	388	0	14451	0.36	44.6	33.7	7.9	13.8
CIQ & BS & Ownerships	German Civil Law	17	4718	26.1	3587	0	103190	0.33	22.2	61.2	8.5	8.1
CIQ & BS & Ownerships	French Civil Law	36	2309	12.8	1387	0	53196	0.36	35.9	42.9	8.7	12.8
CIQ & BS & Ownerships	Other	17	107	0.6	63	0	2034	0.37	28.6	51.4	9.0	10.9
Final Sample	All	52	10136	100.0	5522	0	127179	0.32	33.9	39.1	22.1	5.0
Final Sample	English Common Law	21	6820	67.3	3217	0	94684	0.33	41.1	28.9	24.7	5.3
Final Sample	Scandinavian Civil Law	4	319	3.1	208	0	3020	0.35	26.2	49.0	18.2	7.0
Final Sample	German Civil Law	8	1920	18.9	1476	0	20324	0.28	7.7	74.2	14.9	3.3
Final Sample	French Civil Law	19	1077	10.6	621	0	9151	0.35	19.9	62.5	12.4	5.7
Final Sample	Other	0	0	0.0	0	0	0					

The table provides an overview of our sample at various stages of data preparation. The initial stage involves Capital IQ, which we subsequently merge with balance sheet data data from Worldscope, Compustat, and Amadeus, referred to as "CIQ & BS" in the table. This dataset serves as the population for us. Following this, we integrate the merged dataset with debt ownership data from eMAXX and Dealscan, labeled as "CIQ & BS & Ownerships". From this penultimate sample, we exclude observations where debt ownership observed is less than 30% of the firm's balance sheet total debt, and we further exclude observations from countries with fewer than ten distinct firms to arrive at our final sample. Countries with unknown legal origin or those marked as "Socialist" according to Djankov et al. (2007) are categorized as "Other". The columns represent the legal origin, number of countries, number of firms, percentage of firms in each legal origin, number of public firms, number of firms that never had debt, and leverage (total assets/total debt). The subsequent four columns detail outstanding amounts of Arm's length debt (including bonds and commercial papers), term loans, drawn lines of credit, and other types of debt as per Capital IQ.

Table 2.2: Summary Statistics

Country	Num Firms	Num Firm Quarters	% of Public Firms	Total Assets USD (m)	Total Debt USD (m)	Leverage	Arm's length Debt Observed %	Loans Observed %	Lines Observed %	Identified Instruments %	Total Debt Ownership Observed %
Australia	278	1557	56.65	2,592.39	653.70	27.46	4.47	33.79	46.19	84.45	81.74
Bermuda	55	778	68.12	7,454.64	1,636.46	32.30	26.08	40.27	14.57	80.92	70.57
Canada	577	8200	64.76	4,798.94	1,377.12	30.33	30.88	18.20	37.40	86.48	71.91
Cayman Islands	12	193	58.55	1,835.67	517.37	27.18	16.61	42.06	28.05	86.71	79.41
Guernsey	11	66	19.70	1,609.18	635.08	36.32	15.23	47.93	17.30	80.46	70.56
Hong Kong	194	1237	75.59	4,172.63	743.31	24.34	6.59	70.17	3.77	80.53	75.76
India	202	853	83.47	4,312.84	1,473.40	43.52	3.60	67.05	2.94	73.59	71.40
Ireland	55	587	65.25	9,562.39	2,953.42	31.72	25.81	32.01	24.13	81.95	71.18
Israel	14	68	86.76	2,572.92	1,076.65	38.19	6.07	62.65	13.03	81.75	76.81
Jersey	10	59	18.64	3,251.63	582.84	26.46	26.11	18.57	44.67	89.35	75.00
Malaysia	56	388	75.77	1,632.88	570.56	28.95	0.75	63.94	15.21	79.89	79.29
New Zealand	41	273	53.85	993.11	243.41	26.14	1.15	25.38	61.74	88.27	87.47
Pakistan	16	196	91.33	767.02	138.25	27.18	0.00	78.82	0.00	78.82	78.82
Qatar	12	44	97.73	5,813.67	1,397.85	26.07	0.91	67.07	3.95	71.93	71.05
Saudi Arabia	35	525	96.57	7,569.39	3,130.48	40.47	0.59	78.60	1.01	80.19	79.62
Singapore	108	942	43.10	2,349.95	642.10	29.24	8.80	66.53	4.16	79.49	74.15
South Africa	32	168	83.93	3,914.85	1,072.61	24.30	1.97	50.74	24.16	76.88	75.13
Thailand	51	432	80.79	1,244.22	495.78	37.91	2.12	75.11	0.37	77.60	75.77
United Arab Emirates	33	370	75.41	6,079.63	1,181.99	26.81	7.74	75.38	1.39	84.52	78.49
United Kingdom	579	4347	50.59	4,967.38	962.29	28.99	10.90	34.68	41.05	86.63	80.12
United States	4449	73401	56.68	8,077.90	2,263.36	33.57	41.69	23.62	22.27	87.57	71.45
English Common Law	**6820**	**94684**	**58.19**	**7,286.66**	**2,022.68**	**32.83**	**36.26**	**26.61**	**23.93**	**86.81**	**72.35**
Denmark	31	224	65.18	5,753.88	1,214.20	27.44	8.89	55.55	20.41	84.85	78.77
Finland	59	576	60.59	3,881.11	1,029.62	33.90	17.99	54.16	6.96	79.11	64.13
Norway	131	1334	76.39	2,227.01	662.81	37.93	31.66	35.18	16.55	83.38	65.57
Sweden	98	886	74.72	2,943.71	838.11	33.78	13.94	44.77	23.46	82.17	74.10
Scandinavian Civil Law	**319**	**3020**	**72.05**	**3,014.35**	**825.10**	**35.17**	**22.16**	**43.12**	**17.03**	**82.32**	**68.78**
Austria	22	180	90.00	5,019.25	1,439.35	29.79	18.56	35.66	23.14	77.37	61.06
China	140	889	89.20	3,670.39	771.10	26.59	11.07	69.56	1.18	81.81	74.59
Germany	209	1963	78.76	4,157.44	1,318.02	31.64	7.96	56.69	16.48	81.13	74.98
Japan	997	10990	84.20	2,287.67	602.56	25.65	1.68	65.74	10.64	78.07	76.66
Poland	31	305	59.67	3,144.47	713.42	29.11	8.45	51.03	18.94	78.42	73.49
South Korea	41	220	93.64	4,012.44	1,304.75	31.87	7.44	63.50	5.51	76.45	72.22
Switzerland	68	353	70.82	4,181.25	1,328.50	28.30	9.79	41.22	26.64	77.64	70.34
Taiwan	412	5424	88.57	1,116.89	360.40	32.46	1.20	60.13	16.31	77.64	76.57
German Civil Law	**1920**	**20324**	**84.61**	**2,304.90**	**643.69**	**28.29**	**3.03**	**62.60**	**12.76**	**78.39**	**76.04**
Argentina	19	158	92.41	2,031.68	410.68	23.49	33.15	52.36	0.00	85.52	65.90
Belgium	40	316	61.71	4,156.26	1,390.78	32.37	18.16	47.31	15.97	81.44	68.70
Brazil	41	201	60.20	2,928.90	857.41	30.52	18.38	51.78	2.24	72.41	61.55
Chile	25	203	96.55	4,868.53	1,460.57	29.29	32.24	43.46	0.00	75.70	58.70
Colombia	10	76	98.68	11,538.62	3,347.42	40.59	49.36	28.51	0.26	78.12	52.60
France	270	1828	53.99	5,900.08	1,767.98	32.22	15.62	49.12	16.53	81.27	70.90
Greece	48	280	54.64	1,489.66	725.13	48.43	4.16	63.31	10.94	78.41	75.88
Indonesia	63	781	96.54	1,200.38	421.38	36.77	5.96	62.17	14.34	82.47	78.82
Italy	127	1048	55.44	4,020.21	1,456.70	32.94	12.25	58.77	8.31	79.32	70.31
Kuwait	19	155	74.84	3,103.55	933.04	33.07	0.00	70.34	6.90	77.24	77.24
Luxembourg	26	293	82.25	18,559.37	4,108.73	33.28	29.94	45.58	12.86	88.37	70.97
Mexico	62	938	84.65	6,586.55	2,620.79	34.91	35.50	44.32	2.20	82.02	62.50
Netherlands	97	758	52.11	5,669.77	1,724.40	29.60	12.47	42.52	28.38	83.37	76.35
Oman	13	158	97.47	5,381.85	624.47	32.69	10.24	74.18	1.13	85.55	76.14
Peru	17	198	84.85	1,750.50	443.17	32.71	36.41	44.60	0.00	81.01	60.09
Philippines	25	206	83.50	2,827.69	1,217.17	36.44	9.48	64.08	1.12	74.69	66.49
Russia	49	348	72.13	16,259.83	4,900.19	36.86	0.55	72.49	0.50	73.54	73.04
Spain	102	872	66.17	5,937.51	2,121.28	42.54	9.03	68.24	5.03	82.30	75.76
Turkey	24	334	90.42	4,461.52	1,615.26	40.31	9.61	68.00	0.00	77.61	71.00
French Civil Law	**1077**	**9151**	**69.70**	**5,566.80**	**1,766.34**	**34.68**	**15.93**	**54.81**	**10.06**	**80.80**	**70.88**
World	**10136**	**127179**	**63.57**	**6,265.34**	**1,755.43**	**32.29**	**29.15**	**34.79**	**20.99**	**84.93**	**72.75**

This table presents an overview of the debt structure and coverage within our sample of debt instruments at the firm level, categorized by country. The reported figures represent averages derived from firm-quarter observations within each country. The initial five columns delineate key attributes of firms within our sample for each country, including the number of firms, number of quarterly observations, the percentage of public firms, total assets (in USD millions), total debt (in USD million), and the leverage (total debt/total assets) expressed as a percentage. The subsequent columns provide distinct measures of debt coverage within our sample, all presented as percentages of total debt. These include the outstanding amounts of Arm's length debt specific to instruments in our sample, outstanding amounts of term loans specific to loans in our sample, outstanding amounts of drawn lines of credit specific to lines in our sample, the total outstanding amount of observed debt instruments (comprising bonds, loans, and lines of credit), the total ownership observed for these debt instruments. Missing data are disregarded in calculating the averages.

Table 2.3: Debt Ownership Concentration - Bank Debt

Country	Num of Tranches (loans & lines)	Num of Loans	Num of Lines of Credit	Num Lenders	Num Lead Arrangers	Herfindahl of Lenders Amounts	Num to Cross 10%	Num to Cross 20%	Num to Cross 30%
Australia	3.40	1.34	2.06	8.41	4.77	27.99	1.09	1.43	1.93
Bermuda	3.61	2.66	0.96	10.64	7.40	20.94	1.15	1.74	2.51
Canada	1.91	0.75	1.15	6.92	2.01	24.35	1.12	1.63	2.23
Cayman Islands	4.13	3.55	0.58	8.92	3.25	24.49	1.04	1.36	1.81
Guernsey	1.98	1.35	0.64	4.48	2.80	66.27	1.08	1.23	1.35
Hong Kong	2.21	1.92	0.29	15.30	8.05	16.51	1.18	1.86	2.69
India	3.90	3.67	0.24	14.58	4.53	28.04	1.11	1.49	2.07
Ireland	2.80	1.86	0.95	11.84	4.29	22.67	1.24	1.85	2.52
Israel	2.03	1.62	0.41	7.97	3.41	34.05	1.18	1.61	1.94
Jersey	1.53	0.42	1.10	3.78	3.27	53.07	1.14	1.34	1.62
Malaysia	3.50	3.16	0.34	7.31	4.65	33.73	1.02	1.25	1.64
New Zealand	3.52	0.73	2.80	3.66	2.29	37.43	1.00	1.04	1.27
Pakistan	1.11	1.11	0.00	5.55	3.39	31.54	1.04	1.40	1.88
Qatar	2.20	1.98	0.23	14.86	8.86	25.16	1.34	2.11	3.09
Saudi Arabia	3.85	3.80	0.06	13.02	11.19	27.54	1.15	1.66	2.30
Singapore	2.11	1.85	0.25	7.83	4.49	28.98	1.04	1.36	1.87
South Africa	3.44	2.32	1.12	11.20	6.46	25.04	1.21	1.73	2.40
Thailand	2.40	2.25	0.15	5.94	3.31	35.17	1.01	1.23	1.46
United Arab Emirates	2.38	2.22	0.15	11.12	6.18	20.82	1.26	1.88	2.72
United Kingdom	2.47	1.35	1.11	7.74	4.67	30.48	1.11	1.53	2.03
United States	1.99	1.05	0.95	9.33	2.76	24.49	1.20	1.83	2.54
English Common Law	**2.10**	**1.14**	**0.96**	**9.15**	**3.05**	**25.05**	**1.18**	**1.76**	**2.43**
Denmark	3.19	2.55	0.63	9.12	7.44	24.52	1.29	1.87	2.55
Finland	2.69	2.12	0.57	6.68	4.72	26.01	1.10	1.42	1.82
Norway	2.19	1.59	0.61	4.97	3.54	29.13	1.07	1.38	1.80
Sweden	3.32	2.50	0.81	6.57	5.23	36.10	1.13	1.50	1.91
Scandinavian Civil Law	**2.69**	**2.03**	**0.66**	**6.07**	**4.55**	**30.22**	**1.11**	**1.47**	**1.90**
Austria	5.24	4.54	0.69	9.96	7.01	21.47	1.02	1.33	1.91
China	1.54	1.48	0.06	8.56	4.27	26.98	1.07	1.44	1.96
Germany	4.74	3.99	0.76	12.08	6.90	25.94	1.24	1.84	2.51
Japan	3.89	3.47	0.42	7.74	1.82	26.87	1.02	1.24	1.63
Poland	2.33	1.62	0.71	9.69	7.87	19.17	1.23	1.96	2.90
South Korea	4.58	4.34	0.25	11.20	5.39	22.61	1.00	1.21	1.72
Switzerland	2.45	1.68	0.77	13.41	6.14	23.62	1.32	2.07	2.92
Taiwan	3.70	2.81	0.88	13.16	5.95	13.70	1.11	1.72	2.46
German Civil Law	**3.79**	**3.22**	**0.57**	**9.82**	**3.77**	**22.99**	**1.08**	**1.46**	**2.00**
Argentina	1.24	1.24	0.00	4.13	2.30	35.95	1.07	1.53	2.01
Belgium	2.63	1.73	0.90	10.69	6.06	25.77	1.27	1.89	2.61
Brazil	2.05	1.94	0.11	8.33	3.12	24.12	1.09	1.56	2.09
Chile	1.57	1.55	0.02	9.13	2.74	17.27	1.26	1.98	2.97
Colombia	1.55	1.41	0.14	4.93	2.97	30.07	1.08	1.60	2.22
France	4.00	2.81	1.20	11.82	6.46	18.70	1.15	1.78	2.53
Greece	3.77	3.28	0.49	9.49	4.31	30.32	1.10	1.48	2.01
Indonesia	2.56	1.91	0.65	11.23	5.91	23.13	1.10	1.57	2.14
Italy	3.56	2.70	0.86	10.72	6.21	20.67	1.15	1.74	2.47
Kuwait	2.93	2.66	0.27	15.37	5.57	22.21	1.61	2.47	3.48
Luxembourg	3.13	2.18	0.95	14.35	8.88	27.02	1.23	1.77	2.43
Mexico	2.01	1.78	0.23	10.44	3.64	20.60	1.21	1.81	2.58
Netherlands	3.39	2.34	1.06	11.63	6.34	19.33	1.32	2.06	2.84
Oman	1.87	1.83	0.04	10.08	7.03	19.71	1.23	1.87	2.57
Peru	0.84	0.84	0.00	4.88	1.38	27.47	1.15	1.45	1.90
Philippines	3.70	3.63	0.07	17.90	7.52	24.60	1.17	1.77	2.65
Russia	4.37	4.28	0.09	18.90	11.59	24.85	1.23	1.98	2.78
Spain	4.08	3.45	0.63	19.21	10.38	17.91	1.24	2.00	2.90
Turkey	4.46	4.46	0.00	11.41	6.98	22.10	1.17	1.70	2.47
French Civil Law	**3.23**	**2.58**	**0.65**	**12.11**	**6.32**	**21.47**	**1.19**	**1.81**	**2.55**
World	**2.46**	**1.60**	**0.87**	**9.40**	**3.43**	**24.49**	**1.16**	**1.70**	**2.35**

This table outlines the consolidated ownership concentration measures of bank debt (loans and lines of credit) at the firm level. The reported figures represent averages derived from firm-quarter observations within each country. The defined statistics include the total number of tranches (comprising loans and lines), the count of lenders, the count of domestic lenders, the Herfindahl–Hirschman index of lenders' amount, lending percentages of the 1, 3, 5, and 10 largest lenders in relation to the total tranches amount, the count of lenders to cross the thresholds of 10%, 20%, and 30% of ownership in total tranches amount. Missing data are disregarded in calculating the averages.

Table 2.4: Debt Ownership Concentration - Arm's Length Debt

Country	Bonds Avg size USD m	Bonds Outstanding PAR % of Total Debt	Bond Ownership Observed % of Outstanding PAR	Num of Bond Issues	Num of Bondholders	Herfindahl of Bondholders Ownership	Num to Cross 10%	Num to Cross 20%	Num to Cross 30%
Australia	262.49	4.47	31.23	0.31	7.36	9.16	1.33	2.25	3.73
Bermuda	303.21	26.08	48.64	1.39	69.81	6.69	1.36	2.66	4.45
Canada	252.99	30.88	45.90	2.34	105.21	6.51	1.74	3.53	6.15
Cayman Islands	312.99	16.61	53.91	0.73	38.74	7.44	1.17	2.02	3.30
Guernsey	266.28	15.23	33.46	0.52	23.44	9.68	1.29	2.12	2.88
Hong Kong	326.75	6.59	22.22	0.35	8.52	14.96	1.09	1.86	2.81
India	194.00	3.60	31.91	0.45	7.54	26.58	1.10	1.52	2.21
Ireland	446.12	25.81	58.15	3.54	175.51	4.05	1.86	3.94	6.92
Israel	612.49	6.07	16.72	0.22	6.62	15.79	1.50	2.00	2.50
Jersey	388.48	26.11	42.24	0.73	26.69	11.56	1.06	1.78	2.61
Malaysia	74.37	0.75	10.43	0.06	0.27	46.13	1.00	1.00	1.00
New Zealand	259.38	1.15	11.08	0.05	0.65	3.94	2.00	3.00	5.00
Pakistan	.	0.00	.	0.00	0.00	.	.	.	.
Qatar	420.00	0.91	3.78	0.05	0.23	.	.	.	.
Saudi Arabia	625.00	0.59	1.94	0.01	0.03	.	.	.	.
Singapore	192.53	8.80	32.36	0.64	19.63	11.62	1.36	2.31	3.78
South Africa	172.82	1.97	21.42	0.21	3.96	30.65	1.00	1.00	1.00
Thailand	80.82	2.12	25.66	0.22	0.78	60.37	1.00	1.00	1.00
United Arab Emirates	480.81	7.74	20.30	0.54	16.60	8.98	1.70	3.10	4.85
United Kingdom	385.58	10.90	37.13	0.64	33.72	5.73	1.65	3.38	5.80
United States	350.78	41.69	60.50	3.56	194.08	4.83	1.83	3.86	6.77
English Common Law	**342.00**	**36.26**	**58.23**	**3.05**	**163.47**	**5.05**	**1.82**	**3.81**	**6.67**
Denmark	613.41	8.89	32.70	0.33	17.81	8.05	1.00	1.85	2.92
Finland	267.02	17.99	15.77	0.95	22.96	8.76	1.29	2.50	4.29
Norway	116.24	31.66	38.39	2.69	24.76	10.55	1.22	2.19	3.35
Sweden	197.60	13.94	37.67	0.85	19.17	10.59	1.31	2.24	3.55
Scandinavian Civil Law	**179.52**	**22.16**	**33.54**	**1.64**	**22.26**	**10.44**	**1.24**	**2.20**	**3.40**
Austria	368.49	18.56	11.58	1.12	32.48	5.65	1.38	2.75	5.00
China	287.44	11.07	28.92	0.52	20.00	11.59	1.35	2.29	3.62
Germany	516.02	7.96	19.21	0.38	20.80	6.50	1.34	2.46	4.19
Japan	166.45	1.68	16.94	0.21	2.99	46.16	1.02	1.17	1.44
Poland	102.55	8.45	38.75	0.88	9.29	12.94	1.15	1.96	2.84
South Korea	227.20	7.44	28.02	0.38	8.98	24.17	1.29	1.82	2.71
Switzerland	338.93	9.79	19.63	0.44	12.80	8.12	1.58	3.00	5.16
Taiwan	109.30	1.20	15.07	0.07	0.45	74.96	1.00	1.06	1.18
German Civil Law	**241.51**	**3.03**	**19.03**	**0.23**	**5.37**	**31.01**	**1.17**	**1.76**	**2.61**
Argentina	160.86	33.15	39.58	1.15	28.58	11.09	1.02	1.33	2.14
Belgium	238.64	18.16	22.68	0.97	32.28	8.39	1.11	2.16	3.36
Brazil	235.15	18.38	36.13	0.97	50.18	8.85	1.50	2.67	4.11
Chile	1,016.41	32.24	47.01	1.60	49.23	14.15	1.10	1.79	2.78
Colombia	552.90	49.36	45.96	2.26	213.07	4.48	1.68	3.28	5.60
France	517.06	15.62	29.56	1.16	66.58	9.00	1.67	3.37	5.94
Greece	307.09	4.16	32.25	0.23	7.00	12.92	1.00	1.19	1.54
Indonesia	313.30	5.96	31.75	0.18	8.78	8.24	1.26	2.02	3.33
Italy	552.70	12.25	23.89	0.55	33.61	7.37	1.25	2.19	3.81
Kuwait	.	0.00	.	0.00	0.00	.	.	.	.
Luxembourg	582.39	29.94	41.52	3.09	237.10	3.48	2.44	5.88	10.95
Mexico	269.64	35.50	42.31	3.61	132.37	8.18	1.48	2.84	4.91
Netherlands	582.20	12.47	41.27	0.81	58.53	3.32	2.05	4.41	7.80
Oman	328.85	10.24	5.43	0.41	6.04	.	.	.	.
Peru	155.99	36.41	41.05	1.19	30.53	10.02	1.11	1.72	2.89
Philippines	473.23	9.48	15.91	0.44	10.24	12.51	1.00	1.17	1.50
Russia	171.76	0.55	14.92	0.14	1.36	67.19	1.00	1.00	1.00
Spain	502.74	9.03	22.29	0.36	19.49	6.10	1.33	2.42	3.98
Turkey	339.53	9.61	29.42	0.94	35.60	5.37	1.34	2.61	4.32
French Civil Law	**426.34**	**15.93**	**33.09**	**1.10**	**53.90**	**8.16**	**1.52**	**2.94**	**5.10**
World	**338.51**	**29.15**	**54.86**	**2.43**	**126.97**	**5.51**	**1.79**	**3.73**	**6.53**

This table outlines the consolidated ownership concentration measures of Arm's length debt (bond and commercial papers) at the firm level. The reported figures represent averages derived from firm-quarter observations within each country. The defined statistics include the average size of bonds (in USD millions), outstanding bond amount as a percentage of total debt, bond holding observed in eMAXX as a percentage of bonds outstanding amount (averaged across bonds' par held in eMAXX and weighted by the par outstanding of existing bonds within that quarter), the number of bonds and bondholders, Herfindahl–Hirschman index of bondholder's ownership, bond ownership of the 1, 3, 5, and 10 largest bondholders as percentage observed par ownerships. Missing data are disregarded in calculating the averages.

Table 2.5: Debt Ownership Concentration - All Creditors

Country	Total Debt Ownership Observed %	Num Creditors	Herfindahl of Creditors Ownership	C1	C3	C5	C10	Num to Cross 10%	Num to Cross 20%	Num to Cross 30%
Australia	81.74	15.77	27.20	31.68	55.60	77.35	92.02	1.11	1.47	2.03
Bermuda	70.57	80.44	16.34	22.30	35.65	52.75	74.58	1.26	2.18	3.34
Canada	71.91	112.13	18.84	22.77	40.35	60.27	78.19	1.36	2.30	3.57
Cayman Islands	79.41	47.66	19.03	27.25	45.27	64.14	82.79	1.05	1.54	2.17
Guernsey	70.56	27.92	56.38	61.53	76.49	85.37	91.23	1.14	1.35	1.62
Hong Kong	75.76	23.82	16.07	21.55	37.60	58.17	82.29	1.18	1.89	2.76
India	71.40	22.13	27.88	36.60	51.56	68.33	86.28	1.12	1.50	2.11
Ireland	71.18	187.35	16.21	20.58	34.74	51.93	70.71	1.57	2.80	4.45
Israel	76.81	14.59	33.39	40.49	61.65	76.37	89.62	1.19	1.63	1.99
Jersey	75.00	30.47	42.51	47.96	64.65	77.02	88.24	1.17	1.46	1.83
Malaysia	79.29	7.59	33.74	38.53	64.32	80.93	95.22	1.02	1.24	1.64
New Zealand	87.47	4.32	37.11	42.77	74.58	94.95	99.55	1.00	1.04	1.29
Pakistan	78.82	5.55	31.54	34.31	52.41	78.65	98.90	1.04	1.40	1.88
Qatar	71.05	15.09	25.15	28.11	40.42	62.92	82.79	1.34	2.11	3.09
Saudi Arabia	79.62	13.05	27.53	34.81	48.59	65.24	86.02	1.15	1.67	2.31
Singapore	74.15	27.46	27.74	32.39	54.55	76.86	92.30	1.09	1.48	2.08
South Africa	75.13	15.16	24.95	28.94	48.81	69.19	86.44	1.21	1.73	2.42
Thailand	75.77	6.72	34.78	42.56	67.36	85.66	96.84	1.01	1.22	1.46
United Arab Emirates	78.49	27.72	20.04	23.80	39.30	62.90	85.55	1.29	1.94	2.86
United Kingdom	80.12	41.46	28.22	31.81	51.46	72.28	88.13	1.18	1.75	2.47
United States	71.45	203.41	16.20	21.33	33.19	47.95	65.57	1.59	2.98	4.87
English Common Law	**72.35**	**172.62**	**17.79**	**22.79**	**36.16**	**52.09**	**69.66**	**1.51**	**2.75**	**4.42**
Denmark	78.77	26.93	23.25	26.57	45.60	62.65	85.94	1.29	1.96	2.63
Finland	64.13	29.63	23.88	28.02	51.24	75.58	90.56	1.12	1.54	2.05
Norway	65.57	29.73	21.68	27.57	48.80	69.15	87.17	1.12	1.67	2.35
Sweden	74.10	25.75	30.01	35.33	55.48	74.87	89.35	1.17	1.62	2.19
Scandinavian Civil Law	**68.78**	**28.34**	**24.66**	**29.85**	**50.98**	**71.57**	**88.36**	**1.14**	**1.65**	**2.27**
Austria	61.06	42.44	19.41	26.45	45.63	68.35	89.17	1.03	1.40	2.01
China	74.59	28.57	25.42	32.35	51.30	71.56	89.13	1.10	1.55	2.19
Germany	74.98	32.88	24.70	29.13	47.43	65.98	85.01	1.25	1.87	2.62
Japan	76.66	10.73	26.72	32.84	57.70	82.14	96.38	1.02	1.24	1.64
Poland	73.49	18.98	17.78	20.58	37.54	64.63	84.46	1.25	1.98	2.99
South Korea	72.22	20.18	21.57	29.52	50.38	74.24	91.47	1.03	1.29	1.82
Switzerland	70.34	26.21	21.99	24.96	41.86	58.88	80.50	1.36	2.18	3.10
Taiwan	76.57	13.60	13.75	20.65	35.77	57.94	87.55	1.11	1.72	2.47
German Civil Law	**76.04**	**15.19**	**22.67**	**28.79**	**49.81**	**72.79**	**92.04**	**1.08**	**1.47**	**2.03**
Argentina	65.90	32.70	27.10	31.91	48.75	65.75	87.69	1.06	1.51	2.12
Belgium	68.70	42.97	23.88	27.96	44.56	64.45	83.15	1.24	1.87	2.71
Brazil	61.55	58.51	20.04	24.85	43.24	67.41	86.54	1.12	1.70	2.41
Chile	58.70	58.36	15.38	20.73	36.40	51.86	74.04	1.34	2.21	3.30
Colombia	52.60	218.00	12.20	18.24	29.35	45.65	66.84	1.43	2.41	3.75
France	70.90	78.41	17.14	21.00	37.48	60.77	83.09	1.25	2.06	3.09
Greece	75.88	16.50	29.67	35.56	51.57	68.94	89.77	1.11	1.52	2.05
Indonesia	78.82	20.02	22.02	26.99	48.27	68.82	87.70	1.12	1.64	2.26
Italy	70.31	44.33	18.96	23.08	40.99	63.21	85.29	1.19	1.84	2.66
Kuwait	77.24	15.37	22.21	28.21	47.06	62.82	79.93	1.61	2.46	3.53
Luxembourg	70.97	251.44	17.91	23.38	37.96	54.68	70.06	1.59	2.92	4.80
Mexico	62.50	142.80	13.53	18.90	32.53	50.70	69.92	1.43	2.47	3.92
Netherlands	76.35	70.17	17.52	21.49	37.97	57.59	78.46	1.42	2.34	3.38
Oman	76.14	16.13	19.40	26.43	41.44	61.35	84.24	1.28	1.93	2.70
Peru	60.09	35.40	19.71	26.90	46.03	62.97	81.27	1.15	1.61	2.33
Philippines	66.49	28.15	24.09	29.29	43.21	61.61	82.11	1.17	1.87	2.73
Russia	73.04	20.26	24.78	30.74	45.22	59.39	77.66	1.23	1.96	2.78
Spain	75.76	38.70	17.09	22.56	36.86	55.23	78.08	1.29	2.10	3.12
Turkey	71.00	47.01	18.71	26.16	39.54	59.89	83.45	1.25	1.82	2.61
French Civil Law	**70.88**	**66.00**	**18.90**	**23.82**	**40.11**	**60.03**	**80.84**	**1.27**	**2.04**	**3.02**
World	**72.75**	**136.36**	**18.81**	**23.99**	**38.98**	**56.43**	**74.49**	**1.42**	**2.47**	**3.89**

This table outlines the consolidated ownership concentration measures for all creditor. The reported figures represent averages derived from firm-quarter observations within each country. The reported columns show, respectively, the total ownership observed of identified debt instruments as a percentage of total debt, number of all creditors (holding bonds, loans, or lines of credit), Herfindahl–Hirschman index of creditors's ownership, the ownership stake of the largest 1, 3, 5, and 10 creditors as a percentage of total debt, Number of creditors to cross the 10%, 20%, and 30% of total debt in ownership. Missing data are disregarded in calculating the averages.

Table 2.6: Debt Maturity

Country	Total Debt	Arm's Length Debt	Bank Debt		
			Bank Loans	Credit Lines	Total Bank Debt
Australia	2.55	5.15	2.83	2.10	2.39
Bermuda	4.57	5.49	3.42	2.83	3.26
Canada	4.35	6.87	3.13	2.39	2.58
Cayman Islands	3.26	4.91	2.65	2.64	2.66
Guernsey	4.91	4.38	5.41	2.91	4.89
Hong Kong	2.31	3.46	2.23	2.31	2.18
India	5.69	5.17	5.87	3.20	5.70
Ireland	5.05	8.21	3.34	2.75	3.12
Israel	3.83	8.38	3.70	1.88	3.38
Jersey	3.43	3.64	5.54	2.69	3.47
Malaysia	3.39	3.18	3.84	1.76	3.38
New Zealand	2.37	4.58	2.34	2.33	2.31
Pakistan	5.04	.	5.14	.	5.04
Qatar	5.26	6.37	5.75	1.48	5.20
Saudi Arabia	4.99	4.68	5.04	2.14	4.97
Singapore	2.60	3.55	2.44	2.36	2.39
South Africa	2.30	2.74	2.42	2.30	2.26
Thailand	4.32	3.12	4.43	4.89	4.32
United Arab Emirates	3.65	5.97	3.44	3.30	3.40
United Kingdom	3.50	6.65	3.26	2.64	2.94
United States	5.83	8.45	3.34	2.73	3.01
English Common Law	**5.39**	**8.19**	**3.35**	**2.67**	**3.00**
Denmark	3.48	3.45	3.76	2.94	3.40
Finland	3.26	3.57	3.15	2.80	3.07
Norway	3.32	3.38	3.25	2.64	3.10
Sweden	3.59	3.26	3.83	2.74	3.52
Scandinavian Civil Law	**3.40**	**3.39**	**3.43**	**2.73**	**3.25**
Austria	3.30	2.94	4.06	2.55	3.53
China	3.88	3.56	3.99	2.40	3.87
Germany	3.37	4.44	3.38	2.64	3.23
Japan	3.34	3.77	3.74	1.24	3.36
Poland	3.19	3.24	3.38	2.64	3.19
South Korea	4.05	6.43	3.85	6.84	3.80
Switzerland	3.09	5.01	2.78	2.61	2.72
Taiwan	2.70	2.74	2.84	2.37	2.69
German Civil Law	**3.20**	**3.76**	**3.46**	**1.98**	**3.18**
Argentina	3.02	4.72	2.20	.	2.05
Belgium	4.13	4.82	3.75	2.72	3.44
Brazil	3.34	4.33	3.23	2.06	3.14
Chile	5.39	8.60	2.77	1.61	2.63
Colombia	5.70	7.61	3.69	3.48	3.48
France	3.57	4.60	3.38	2.83	3.19
Greece	4.59	4.97	4.61	4.49	4.64
Indonesia	2.96	3.94	3.07	2.18	2.82
Italy	3.21	4.16	3.06	2.60	3.00
Kuwait	1.91	.	1.99	1.87	1.91
Luxembourg	3.81	5.12	3.55	2.60	3.30
Mexico	5.32	7.99	2.75	2.62	2.73
Netherlands	3.55	6.35	3.16	2.42	2.90
Oman	5.35	1.89	5.40	7.86	5.31
Peru	4.38	6.20	2.82	.	2.68
Philippines	3.75	4.81	3.42	3.07	3.39
Russia	3.27	4.56	3.32	2.30	3.26
Spain	3.70	5.51	3.42	2.85	3.34
Turkey	3.35	4.97	3.35	.	3.33
French Civil Law	**3.77**	**5.67**	**3.27**	**2.70**	**3.14**
World	**4.88**	**7.78**	**3.37**	**2.60**	**3.05**

This table outlines the maturity durations, measured in years, for different types of debt. The reported figures represent averages derived from firm-quarter observations within each country. The columns present the overall maturity of debt, the maturity of Arm's length debt (bonds and commercial papers), and maturity of bank debt (loans and lines of credit). Averages within firm-quarters are computed with weights assigned based on the par value of debt instruments. Missing data are disregarded in calculating the averages.

Table 2.7: Secured Debt

Country	Secured Debt (% of Total Debt)	Arm's Length Secured Debt (% of Total Secured Debt)	Bank Secured Debt (% of Total Secured Debt)
Australia	18.77	3.23	96.77
Bermuda	47.38	1.68	98.32
Canada	27.11	10.48	89.52
Cayman Islands	42.32	28.11	71.89
Guernsey	33.57	0.00	100.00
Hong Kong	13.51	5.40	94.60
India	33.32	3.51	96.49
Ireland	36.03	2.41	97.59
Israel	70.03	0.00	100.00
Jersey	30.61	33.33	66.67
Malaysia	34.75	0.00	100.00
New Zealand	13.77	0.00	100.00
Pakistan	0.01	0.00	100.00
Qatar	39.56	0.00	100.00
Saudi Arabia	47.51	0.00	100.00
Singapore	17.12	3.60	96.40
South Africa	11.73	0.00	100.00
Thailand	19.26	4.07	95.93
United Arab Emirates	27.37	3.64	96.36
United Kingdom	27.10	7.15	92.85
United States	37.60	11.91	88.09
English Common Law	**35.16**	**11.08**	**88.92**
Denmark	43.94	0.00	100.00
Finland	17.80	4.13	95.87
Norway	45.66	12.00	88.00
Sweden	32.77	2.69	97.31
Scandinavian Civil Law	**36.44**	**7.63**	**92.37**
Austria	6.88	0.00	100.00
China	27.32	3.40	96.60
Germany	22.93	3.65	96.35
Japan	2.81	3.80	96.20
Poland	33.21	15.89	84.11
South Korea	23.83	0.00	100.00
Switzerland	26.90	1.60	98.40
Taiwan	58.04	0.13	99.87
German Civil Law	**21.70**	**1.44**	**98.56**
Argentina	17.95	0.00	100.00
Belgium	30.03	2.46	97.54
Brazil	31.42	2.74	97.26
Chile	1.35	50.00	50.00
Colombia	17.47	94.44	5.56
France	21.50	4.42	95.58
Greece	51.20	0.00	100.00
Indonesia	25.25	5.00	95.00
Italy	21.00	6.56	93.44
Kuwait	24.88	0.00	100.00
Luxembourg	39.68	6.25	93.75
Mexico	10.13	36.74	63.26
Netherlands	25.79	9.49	90.51
Oman	30.65	0.00	100.00
Peru	16.03	0.00	100.00
Philippines	15.97	0.00	100.00
Russia	61.56	0.00	100.00
Spain	33.83	5.71	94.29
Turkey	18.53	0.00	100.00
French Civil Law	**24.98**	**6.16**	**93.84**
World	**32.31**	**9.72**	**90.28**

This table describes the composition of debt in terms of security within each country. The reported figures represent averages derived from firm-quarter observations within each country in our sample. The columns present the secured debt (any lien) percentage of total debt, arm's length secured debt percentage of total secured debt, and bank secured debt percentage of total secured debt. Averages within firm-quarters are computed with weights assigned based on the par value of debt instruments. Missing data are disregarded in calculating the averages.

Table 2.8: Debt Currency

Country	Local Currency %	USD %	EUR %	GBP %	JPY %	CAD %	AUD %	NZD %	NOK %	SEK %	CHF %	Other Currencies %	Arm's Length Debt in USD (foreign) % of Total Arm's Length Debt	Bank Debt in USD (foreign) % of Total Bank Debt
Australia	76.35	19.36	1.72	0.71	0.00	0.18	76.35	0.92	0.00	0.00	0.02	0.73	50.61	16.55
Bermuda	0.26	95.16	1.15	1.15	0.00	0.00	0.09	0.00	0.23	0.00	0.01	2.20	92.20	94.90
Canada	61.69	36.97	0.92	0.27	0.00	61.69	0.08	0.00	0.00	0.00	0.00	0.06	42.97	33.28
Cayman Islands	0.00	84.32	0.66	0.02	0.00	0.88	0.00	0.00	0.00	0.00	0.00	14.11	94.37	76.82
Guernsey	0.00	57.65	11.27	31.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	73.13	58.58
Hong Kong	62.48	29.79	0.31	0.51	0.00	0.00	0.09	0.00	0.00	0.00	0.06	69.25	54.29	24.61
India	75.72	21.85	2.10	0.06	0.01	0.02	0.00	0.00	0.00	0.00	0.04	75.92	25.73	13.61
Ireland	25.06	70.80	25.06	3.62	0.00	0.12	0.33	0.00	0.01	0.00	0.01	0.06	87.23	68.39
Israel	17.70	80.47	1.60	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	17.90	65.93	82.00
Jersey	0.00	39.67	10.17	49.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.98	62.50	39.15
Malaysia	69.10	22.08	0.44	0.00	0.00	0.00	2.91	0.11	0.00	0.00	0.00	74.46	7.45	21.25
New Zealand	84.82	7.03	0.10	1.53	0.00	0.00	5.38	84.82	0.00	0.00	0.00	1.14	7.80	6.13
Pakistan	90.70	8.58	0.72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	90.70	0.00	8.59
Qatar	46.21	51.47	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	48.53	22.22	51.80
Saudi Arabia	77.09	22.75	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	77.20	13.79	23.33
Singapore	64.83	30.39	0.20	0.12	0.01	0.00	0.65	0.00	0.00	0.00	0.00	68.63	36.80	29.48
South Africa	50.60	31.11	10.92	4.10	0.00	0.00	1.04	0.00	0.00	0.00	2.09	50.74	33.05	30.71
Thailand	65.18	31.71	0.47	0.00	0.17	0.00	0.55	0.00	0.00	0.00	0.00	67.11	4.15	31.98
United Arab Emirates	43.14	55.00	0.80	0.48	0.00	0.06	0.05	0.00	0.04	0.00	0.21	43.36	65.13	53.47
United Kingdom	68.19	24.48	6.49	68.19	0.00	0.06	0.14	0.00	0.27	0.02	0.04	0.32	46.66	23.72
United States	98.11	98.11	1.29	0.21	0.02	0.22	0.06	0.00	0.00	0.01	0.01	0.06	0.00	0.00
English Common Law	**89.93**	**84.01**	**1.63**	**3.31**	**0.01**	**5.48**	**1.35**	**0.26**	**0.02**	**0.01**	**0.02**	**3.91**	**5.35**	**6.76**
Denmark	34.17	15.62	47.53	2.02	0.00	0.00	0.00	0.00	0.64	0.00	0.00	34.19	0.00	16.98
Finland	98.49	1.14	98.49	0.04	0.00	0.00	0.00	0.00	0.01	0.15	0.10	0.07	0.72	1.74
Norway	50.16	32.62	14.67	0.82	0.00	0.08	0.00	0.00	50.16	1.32	0.01	0.32	29.24	34.09
Sweden	60.64	14.42	23.36	0.29	0.00	0.06	0.00	0.00	0.86	60.64	0.07	0.30	9.94	15.81
Scandinavian Civil Law	**61.09**	**20.11**	**35.34**	**0.61**	**0.00**	**0.05**	**0.00**	**0.00**	**22.57**	**18.49**	**0.04**	**2.79**	**17.29**	**20.92**
Austria	96.39	3.47	96.39	0.04	0.00	0.03	0.00	0.00	0.00	0.00	0.01	0.07	4.46	4.55
China	70.18	25.90	0.13	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	73.93	46.28	18.77
Germany	94.73	4.74	94.73	0.09	0.00	0.13	0.00	0.00	0.00	0.00	0.21	0.10	8.27	5.13
Japan	98.23	1.48	0.26	0.00	98.23	0.00	0.00	0.00	0.00	0.00	0.00	0.03	2.18	0.86
Poland	71.62	9.90	17.45	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.80	71.63	0.68	10.50
South Korea	58.61	40.86	0.33	0.05	0.03	0.00	0.01	0.00	0.00	0.00	0.09	58.64	19.66	33.75
Switzerland	44.29	28.07	24.93	2.65	0.00	0.00	0.00	0.00	0.00	0.03	44.29	0.03	24.66	26.77
Taiwan	85.13	13.92	0.42	0.11	0.02	0.02	0.00	0.00	0.00	0.00	0.00	85.52	6.32	12.00
German Civil Law	**91.40**	**7.22**	**10.72**	**0.08**	**53.31**	**0.02**	**0.00**	**0.00**	**0.00**	**0.00**	**0.80**	**27.84**	**6.57**	**5.96**
Argentina	35.25	63.59	1.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	35.25	70.48	36.32
Belgium	74.11	24.57	74.11	0.31	0.00	0.34	0.19	0.09	0.17	0.00	0.01	0.20	22.99	23.70
Brazil	62.89	32.68	4.35	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	62.96	16.63	35.28
Chile	1.70	43.92	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	55.93	33.22	74.76
Colombia	26.33	73.59	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	26.41	84.83	46.94
France	91.85	7.38	91.85	0.54	0.00	0.09	0.00	0.00	0.01	0.01	0.04	0.09	11.17	7.43
Greece	41.59	58.36	41.59	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.03	62.80	58.50
Indonesia	10.19	89.03	0.58	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	10.36	62.44	88.28
Italy	97.24	2.50	97.24	0.19	0.00	0.00	0.04	0.00	0.00	0.00	0.01	0.01	12.25	1.47
Kuwait	83.66	15.55	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	84.04	1.17	16.89
Luxembourg	34.70	63.56	34.70	1.41	0.00	0.03	0.00	0.00	0.00	0.01	0.11	0.18	48.58	62.80
Mexico	27.26	70.17	1.95	0.21	0.00	0.00	0.00	0.00	0.00	0.00	0.05	27.61	68.88	52.32
Netherlands	70.91	26.22	70.91	2.43	0.01	0.23	0.08	0.00	0.02	0.00	0.00	0.10	44.96	23.54
Oman	60.66	38.84	0.48	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	60.68	33.86	36.98
Peru	11.78	88.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	11.82	76.28	78.27
Philippines	24.08	72.56	1.19	0.48	1.27	0.02	0.01	0.00	0.00	0.00	0.07	24.42	42.63	69.90
Russia	18.43	66.56	14.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	18.46	46.15	64.91
Spain	93.28	5.40	93.28	0.19	0.00	0.14	0.09	0.00	0.00	0.00	0.01	0.89	10.33	5.60
Turkey	10.25	50.51	39.09	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	10.30	62.17	47.72
French Civil Law	**58.40**	**37.07**	**50.32**	**0.44**	**0.03**	**0.06**	**0.03**	**0.00**	**0.01**	**0.00**	**0.02**	**12.02**	**37.87**	**33.98**
World	**87.30**	**66.92**	**7.26**	**2.53**	**8.57**	**4.10**	**1.01**	**0.19**	**0.54**	**0.44**	**0.14**	**8.29**	**7.10**	**8.80**

This table describes the composition of debt in terms of currency of issue within each country. The reported figures represent averages derived from firm-quarter observations within each country. The columns illustrate the percentages of total debt issued in the local currency, USD, EUR, and other major currencies. Averages within firm-quarters are computed with weights assigned based on the par value of debt instruments. Missing data are disregarded in calculating the averages.

Table 2.9: Number of Creditors

	(1) All Creditors	(2) Arm's Length Creditors	(3) Banks	(4) Lead Arrangers	(5) All Creditors	(6) Arm's Length Creditors	(7) Banks	(8) Lead Arrangers	(9) All Creditors	(10) Arm's Length Creditors	(11) Banks	(12) Lead Arrangers
French	-80.78^{***}	-82.90^{***}	2.12^{**}	2.69^{***}								
	(20.67)	(20.87)	(0.99)	(0.70)								
German	-35.57	-38.47^{*}	2.90^{**}	1.00								
	(23.13)	(22.45)	(1.38)	(1.14)								
Scandinavian	-107.41^{***}	-106.06^{***}	-1.35^{***}	1.75^{***}								
	(14.39)	(14.49)	(0.28)	(0.44)								
anti self dealing Index					84.20	88.38	-4.19^{**}	-1.91				
					(72.46)	(72.89)	(1.79)	(2.18)				
SizeGroup=0 × anti self dealing Index									859.04^{***}	863.85^{***}	-4.81	-1.04
									(260.10)	(264.90)	(5.87)	(2.40)
SizeGroup=1 × anti self dealing Index									417.48^{***}	423.82^{***}	-6.34	-2.63
									(118.30)	(121.11)	(3.77)	(2.15)
SizeGroup=2 × anti self dealing Index									137.01^{***}	141.54^{***}	-4.54^{**}	-2.21
									(48.90)	(49.13)	(1.93)	(2.10)
SizeGroup=3 × anti self dealing Index									46.82	50.69	-3.87^{**}	-1.67
									(58.83)	(59.26)	(1.76)	(2.25)
log assets	98.14^{***}	95.74^{***}	2.41^{***}	0.78^{***}	100.29^{***}	98.00^{***}	2.29^{***}	0.72^{***}	150.49^{***}	148.42^{***}	2.07^{***}	0.65^{*}
	(23.97)	(24.18)	(0.25)	(0.27)	(22.62)	(22.77)	(0.20)	(0.23)	(42.16)	(42.71)	(0.61)	(0.33)
log age	4.54	5.34	-0.79^{***}	-0.48^{***}	0.62	1.33	-0.71^{***}	-0.35^{***}	-1.26	-0.55	-0.71^{***}	-0.36^{***}
	(5.40)	(5.44)	(0.13)	(0.16)	(3.55)	(3.58)	(0.13)	(0.11)	(3.14)	(3.15)	(0.12)	(0.11)
Log GDP per capita	20.67^{**}	21.69^{***}	-1.03^{**}	-0.50^{*}	33.26^{***}	34.94^{***}	-1.67^{***}	-0.98^{***}	6.11	7.65	-1.54^{***}	-0.93^{***}
	(7.75)	(7.85)	(0.51)	(0.25)	(7.95)	(7.87)	(0.49)	(0.31)	(16.57)	(16.65)	(0.48)	(0.27)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All	All	All	All
Adj. R2	0.439	0.430	0.146	0.168	0.438	0.430	0.140	0.144	0.505	0.498	0.141	0.146
N	122049	122049	122049	122049	119896	119896	119896	119896	119896	119896	119896	119896

The table investigates the association of ownership dispersion, measured by the number of different types of creditors with the legal environment. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2.10: Creditor Concentration- Number to Cross 30%

	(1) All Creditors	(2) Arm's Length Creditors	(3) Banks	(4) All Creditors	(5) Arm's Length Creditors	(6) Banks	(7) All Creditors	(8) Arm's Length Creditors	(9) Banks
French	-0.98***	-0.92**	0.08						
	(0.25)	(0.45)	(0.08)						
German	-0.65	-2.31**	-0.02						
	(0.43)	(1.13)	(0.26)						
Scandinavian	-1.53***	-0.34	-0.44***						
	(0.17)	(0.48)	(0.04)						
anti self dealing Index				1.07	0.28	-0.14			
				(1.02)	(0.75)	(0.28)			
SizeGroup=0 × anti self dealing Index							7.85***	15.27***	0.39
							(1.89)	(2.53)	(0.52)
SizeGroup=1 × anti self dealing Index							3.35***	4.83***	-0.27
							(0.99)	(0.72)	(0.38)
SizeGroup=2 × anti self dealing Index							1.22	1.35*	-0.16
							(0.82)	(0.69)	(0.31)
SizeGroup=3 × anti self dealing Index							0.96	0.25	-0.11
							(0.94)	(0.71)	(0.28)
log assets	1.38***	2.32***	0.46***	1.41***	2.33***	0.46***	1.76***	2.58***	0.46***
	(0.27)	(0.11)	(0.06)	(0.25)	(0.10)	(0.05)	(0.34)	(0.14)	(0.05)
log age	0.07	-0.19***	-0.04**	0.01	-0.21***	-0.05**	-0.00	-0.20***	-0.05**
	(0.04)	(0.05)	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
Log GDP per capita	0.27**	0.63**	0.01	0.45***	1.19***	-0.01	0.27**	1.09***	-0.00
	(0.12)	(0.29)	(0.05)	(0.12)	(0.16)	(0.06)	(0.12)	(0.14)	(0.05)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All
Adj. R2	0.474	0.543	0.241	0.472	0.541	0.240	0.500	0.550	0.243
N	122049	46841	101452	119896	46474	99494	119896	46474	99494

The table investigates the association of ownership dispersion, measured by the minimum number of different types of creditors to cross the 30% in ownership, in their debt type (all, arm's length debt, or bank debt) with the legal environment. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2.11: Secured Debt

	(1) Total Secured Debt	(2) Secured Arm's Length Debt	(3) Secured Bank Debt	(4) Total Secured Debt	(5) Secured Arm's Length Debt	(6) Secured Bank Debt	(7) Total Secured Debt	(8) Secured Arm's Length Debt	(9) Secured Bank Debt
French	-21.27***	-1.34	-19.79***						
	(3.99)	(0.89)	(3.87)						
German	-34.25***	-2.29***	-30.74***						
	(7.35)	(0.77)	(6.61)						
Scandinavian	-22.65***	-1.64	-22.73***						
	(6.54)	(1.56)	(5.71)						
CR				6.00***	0.13	5.36***			
				(1.60)	(0.21)	(1.51)			
SizeGroup=0 × CR							-13.36***	-0.20	-12.88***
							(2.31)	(0.52)	(2.35)
SizeGroup=1 × CR							1.78	-0.08	1.71
							(2.08)	(0.35)	(2.25)
SizeGroup=2 × CR							4.44***	0.25	3.72***
							(1.36)	(0.26)	(1.27)
SizeGroup=3 × CR							6.79***	0.06	6.19***
							(1.71)	(0.26)	(1.62)
log assets	-10.53***	0.10	-10.46***	-10.51***	0.06	-10.40***	-12.73***	0.05	-12.49***
	(2.11)	(0.16)	(1.97)	(2.12)	(0.14)	(2.00)	(1.49)	(0.16)	(1.40)
log age	-5.48***	0.37	-5.90***	-5.89***	0.31	-6.22***	-5.63***	0.31	-5.97***
	(0.82)	(0.23)	(0.62)	(1.05)	(0.26)	(0.79)	(0.92)	(0.27)	(0.67)
Log GDP per capita	-3.06*	0.91**	-4.18**	-10.53***	0.23	-10.85***	-8.72***	0.26	-9.16***
	(1.81)	(0.39)	(1.63)	(3.39)	(0.38)	(3.12)	(2.79)	(0.38)	(2.54)
Fixed effects	I & Q	I & Q	I & Q	I & Q & R	I & Q & R	I & Q & R	I & Q & R	I & Q & R	I & Q & R
Cluster	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All
Adj. R2	0.285	0.121	0.306	0.282	0.126	0.302	0.294	0.126	0.314
N	121587	121587	121587	120368	120368	120368	120368	120368	120368

The table investigates the association of secured debt, measured as a percentage of total debt type (all debt, arm's length debt, or bank debt) with the legal environment. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category. Additionally, the creditor rights index, which ranges from 0 to 4, is included. A score of 1 is assigned for each of the following rights of secured lenders as defined in laws and regulations: (i) there are restrictions, such as creditor consent or minimum dividends, preventing a debtor from filing for reorganization; (ii) secured creditors can seize their collateral after the reorganization petition is approved (i.e., there is no automatic stay or asset freeze); (iii) secured creditors are paid first from the proceeds of liquidating a bankrupt firm, rather than other creditors such as the government or workers; and (iv) management does not retain administration of the property pending the resolution of the reorganization. The index is derived from Djankov et al. (2007), who extended, revised, and updated the original index compiled by La Porta et al. (1997, 1998). SizeGroup includes the quartile of firms' size (total assets in USD) within the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Columns (4) to (9) also include region fixed effects corresponding to the continents as classified by the World Bank. Controls for the logarithm of GDP per capita for each year in each country, the logarithm of firm age, and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2.12: Debt Maturity

	(1) Total Debt	(2) Arm's Length Debt	(3) Bank Debt	(4) Total Debt	(5) Arm's Length Debt	(6) Bank Debt	(7) Total Debt	(8) Arm's Length Debt	(9) Bank Debt
French	-1.41***	-1.90***	-0.26						
	(0.28)	(0.67)	(0.21)						
German	-0.69***	-3.45***	0.16						
	(0.17)	(0.42)	(0.25)						
Scandinavian	-1.20***	-2.92***	0.30**						
	(0.30)	(0.40)	(0.13)						
anti self dealing Index				1.01	3.50	-0.56*			
				(1.04)	(2.09)	(0.30)			
SizeGroup=0 × anti self dealing Index							5.55***	10.05***	-0.56
							(1.14)	(2.56)	(0.47)
SizeGroup=1 × anti self dealing Index							2.65***	6.11***	-0.58*
							(0.82)	(1.88)	(0.34)
SizeGroup=2 × anti self dealing Index							1.13	4.30**	-0.47
							(0.79)	(1.90)	(0.29)
SizeGroup=3 × anti self dealing Index							0.93	3.41	-0.62*
							(1.03)	(2.04)	(0.32)
log assets	0.97***	1.14***	0.13**	0.99***	1.17***	0.11**	1.24***	1.34***	0.12
	(0.16)	(0.07)	(0.06)	(0.15)	(0.06)	(0.04)	(0.18)	(0.07)	(0.08)
log age	0.41***	0.52***	-0.08*	0.37***	0.45***	-0.05	0.36***	0.46***	-0.05
	(0.06)	(0.04)	(0.04)	(0.07)	(0.07)	(0.05)	(0.07)	(0.06)	(0.04)
Log GDP per capita	-0.04	-0.02	-0.33**	0.24	0.59	-0.27*	0.11	0.51	-0.28*
	(0.11)	(0.40)	(0.16)	(0.19)	(0.39)	(0.16)	(0.17)	(0.39)	(0.16)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All
Adj. R2	0.359	0.306	0.055	0.357	0.290	0.052	0.366	0.292	0.053
N	122049	54028	101452	119896	53407	99494	119896	53407	99494

The table investigates the association of debt maturity, measured in years, with the legal environment. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table 2.13: Debt Currency

	(1) Debt in Local Currency %	(2) Arm's length Debt in Local Currency %	(3) Bank Debt in Local Currency %	(4) Debt in USD (As Foreign) %	(5) Arm's length Debt in USD (foreign)%	(6) Bank Debt in USD (foreign)%	(7) Debt in Local Currency %	(8) Debt in USD (As Foreign) %	(9) Debt in Local Currency %	(10) Debt in USD (As Foreign) %
French	-21.42**	-17.19***	-4.21	20.24**	5.97*	13.76**				
	(9.51)	(4.15)	(6.91)	(8.48)	(3.40)	(5.90)				
German	3.12	-13.45***	17.98***	-1.61	0.14	-1.70				
	(4.08)	(2.73)	(5.06)	(3.57)	(0.82)	(2.79)				
Scandinavian	-30.53***	-10.70**	-21.35**	13.45	2.50	10.80*				
	(11.06)	(4.28)	(8.72)	(8.83)	(2.86)	(5.73)				
anti self dealing Index							0.21	-0.17		
							(19.10)	(15.37)		
SizeGroup=0 × anti self dealing Index									9.02	-7.56
									(19.32)	(15.73)
SizeGroup=1 × anti self dealing Index									10.05	-8.14
									(18.10)	(14.44)
SizeGroup=2 × anti self dealing Index									5.57	-4.61
									(18.20)	(14.47)
SizeGroup=3 × anti self dealing Index									-3.39	2.80
									(19.52)	(15.96)
log assets	-2.39*	8.08***	-11.03***	1.94	1.32	0.54	-2.53	2.06	-0.92	0.73
	(1.38)	(2.17)	(1.14)	(1.35)	(0.92)	(0.43)	(1.60)	(1.53)	(1.71)	(1.73)
log age	2.57*	3.16***	-0.96	-2.37*	-0.24	-2.10	1.67	-1.56	1.70	-1.58
	(1.47)	(0.45)	(1.26)	(1.38)	(0.30)	(1.26)	(1.11)	(1.01)	(1.09)	(1.00)
Log GDP per capita	11.62***	5.88***	5.49	-10.54***	-2.85***	-7.60**	15.75***	-14.97***	14.75***	-14.15***
	(3.95)	(1.46)	(3.43)	(3.65)	(1.06)	(3.10)	(4.75)	(4.40)	(4.81)	(4.43)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All	All
Adj. R2	0.230	0.381	0.379	0.234	0.093	0.194	0.192	0.214	0.197	0.218
N	121587	121587	121587	121692	121692	121692	119448	119553	119448	119553

The table investigates the association of debt currency, with the legal environment. The dependent variables include the percentage of debt in the local currency, and the percentage of debt in USD as a foreign currency (so it is zero for US firms). The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

2.7 Figures

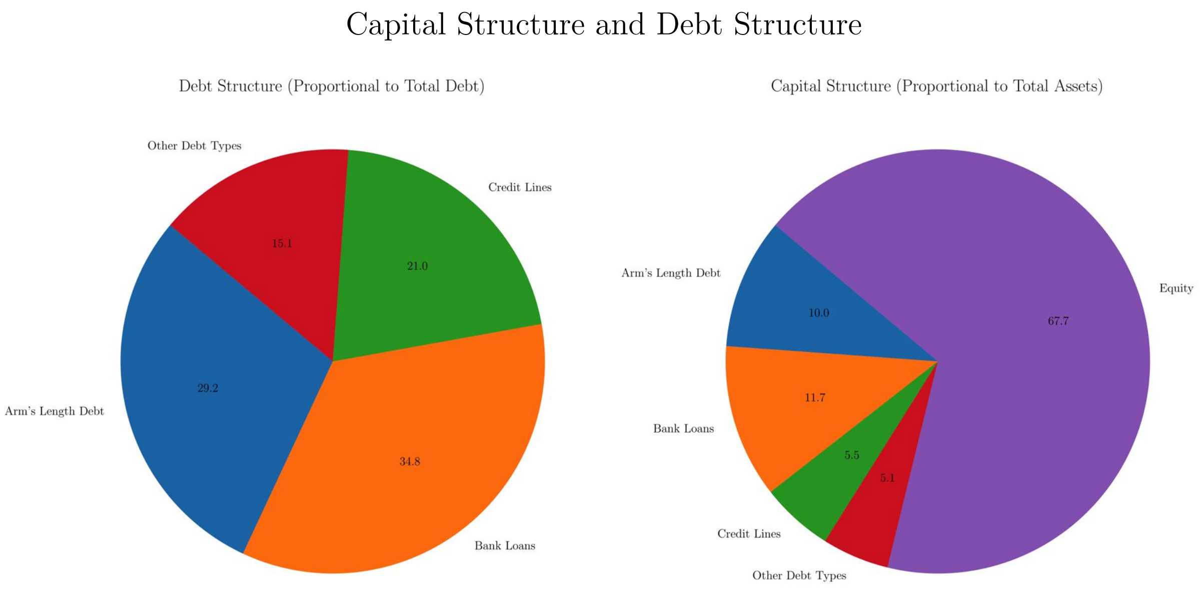


Figure 2.1: This figure shows the global average capital structure and debt structure, including the composition of various debt types. The debt categories comprise arm's-length debt, bank loans, credit lines, and other forms of debt. Detailed variable definitions and sources can be found in Appendix Section A. Averages are computed over all years in the sampleand missing data are disregarded in calculating the averages.

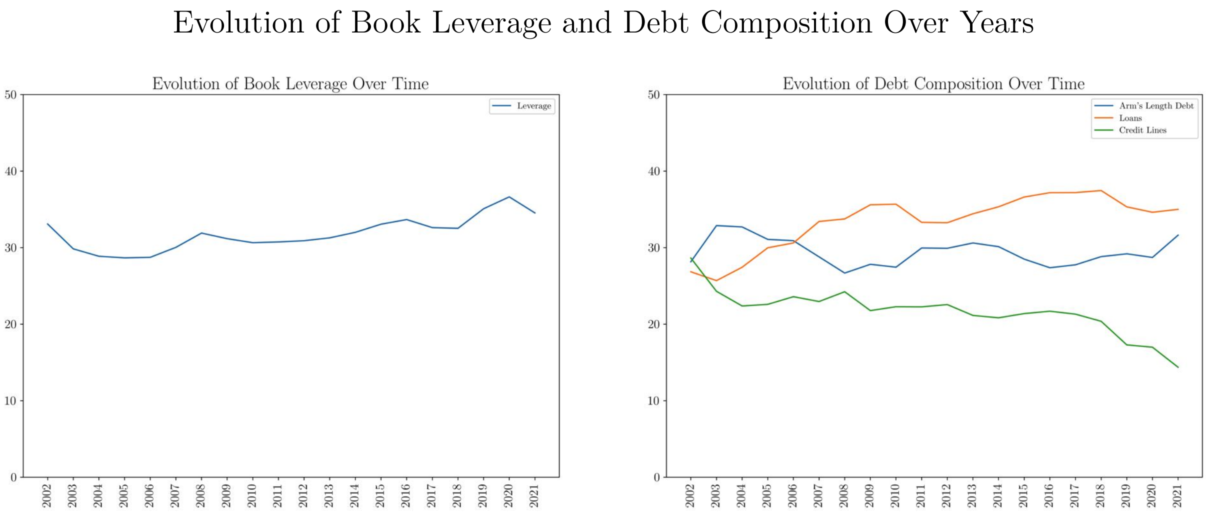


Figure 2.2: This figure displays the global evolution of book leverage and debt composition over years. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

World Map with Legal Origins and the Percentage of Bank Debt and Arm's Length Debt

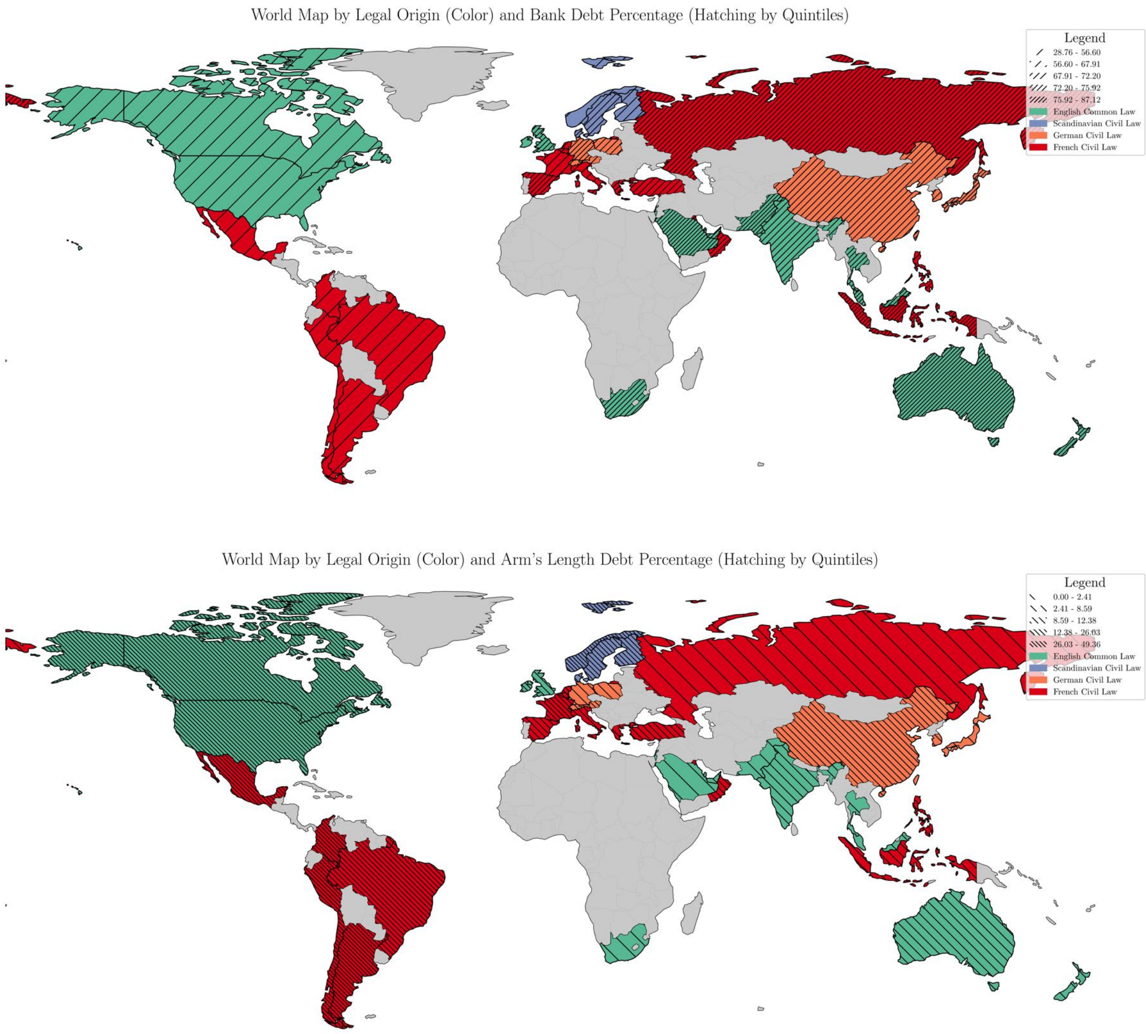


Figure 2.3: This figure presents a world map illustrating legal origins by color and the composition of debt across countries in our sample. The intensity of hatching corresponds to quintiles of debt types, expressed as percentages of total debt. These debt types include the outstanding amounts of arm's-length debt (based on instruments included in our sample) and bank debt (based on loans and credit lines included in our sample). Detailed variable definitions and sources can be found in Appendix Section A. Country-level averages are calculated over all years in the sample and missing data are disregarded in calculating the averages.

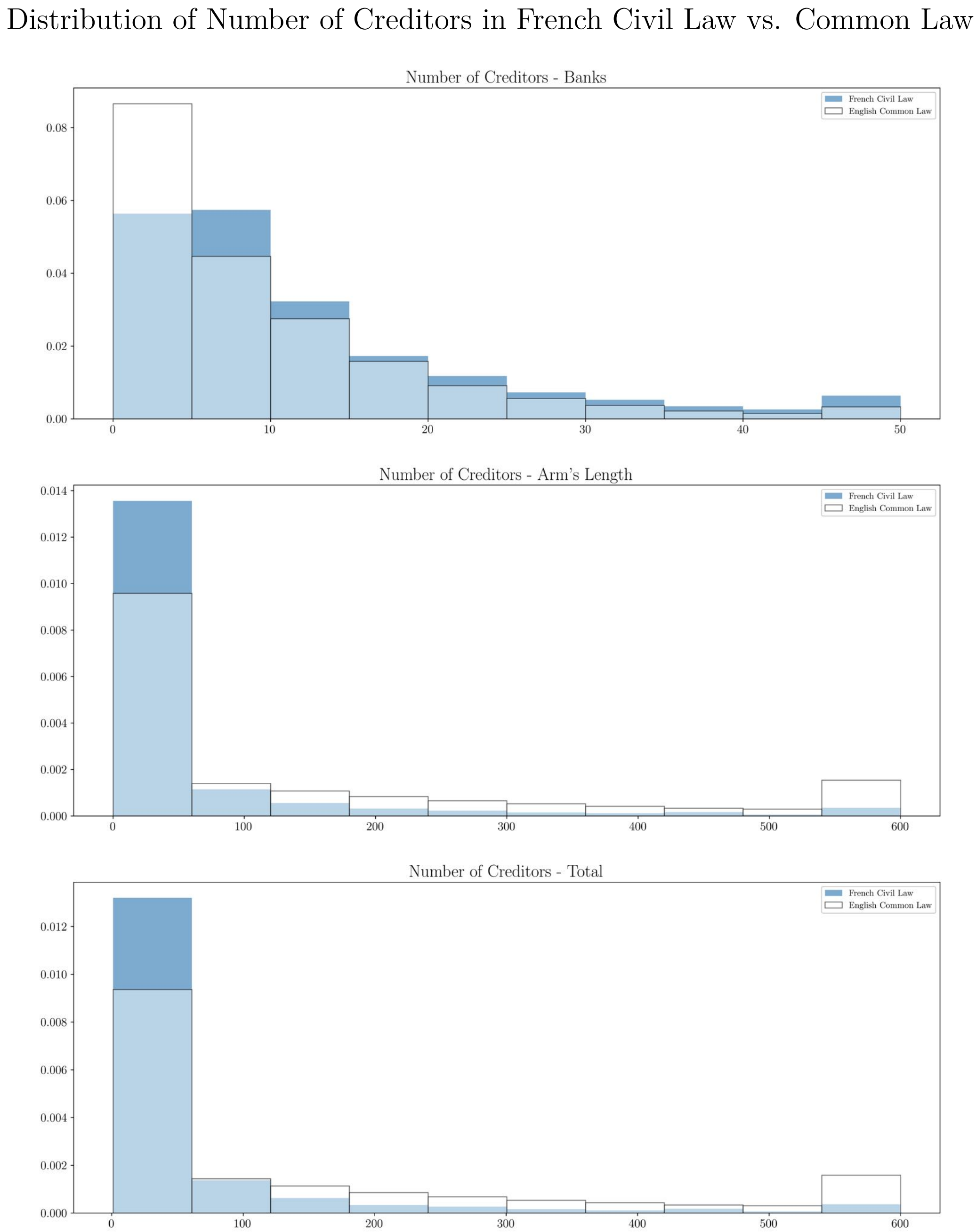


Figure 2.4: This figure displays the distribution of number of creditors, separately for bank debt, arm's length debt, and total debt, comparing French Civil Law countries and English Common Law countries. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

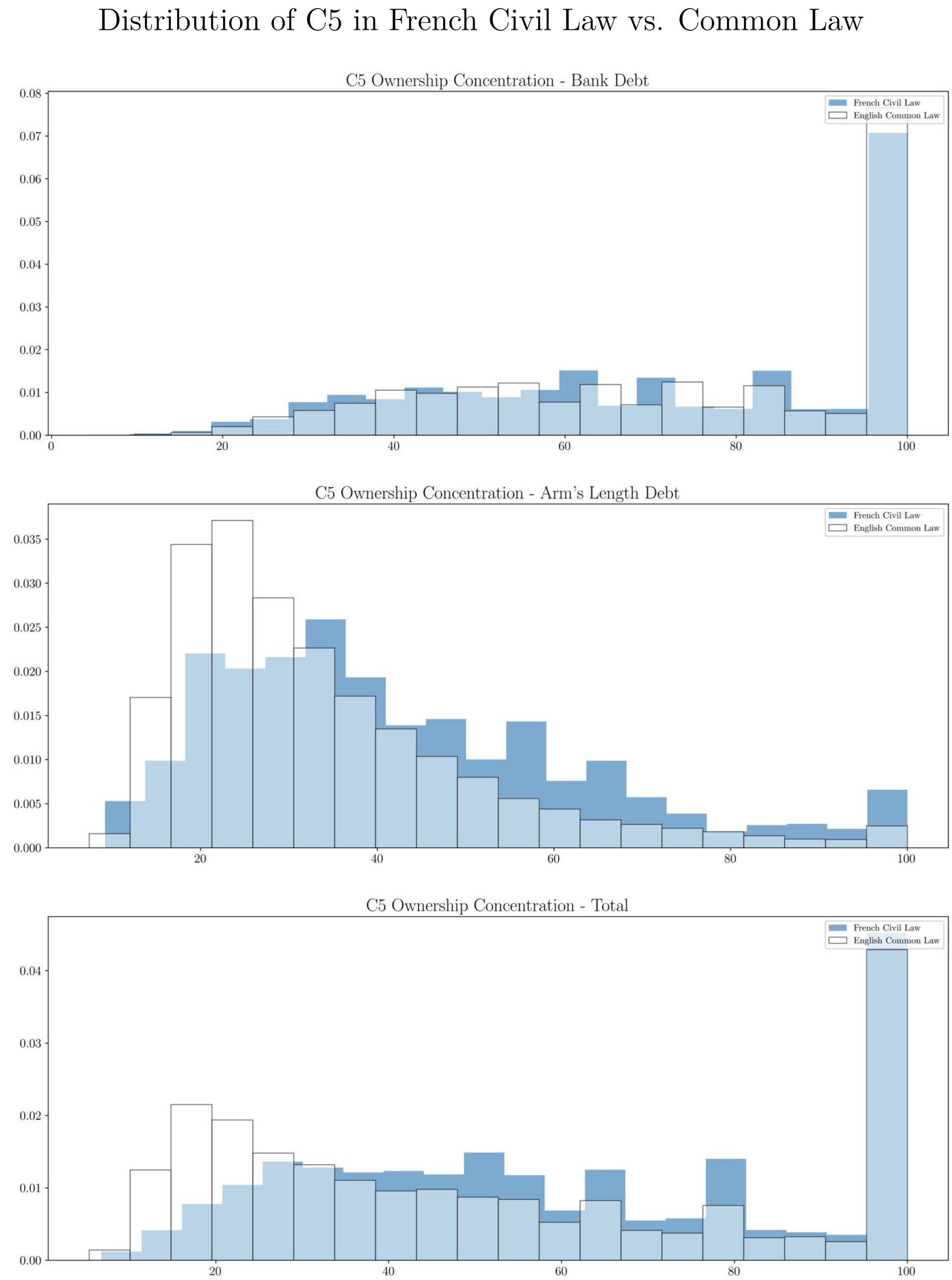


Figure 2.5: This figure displays the distribution of C5, separately for bank debt, arm's length debt, and total debt, comparing French Civil Law countries and English Common Law countries. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

Distribution of NtC 30% of Credit (Types) in French Civil Law vs. Common Law

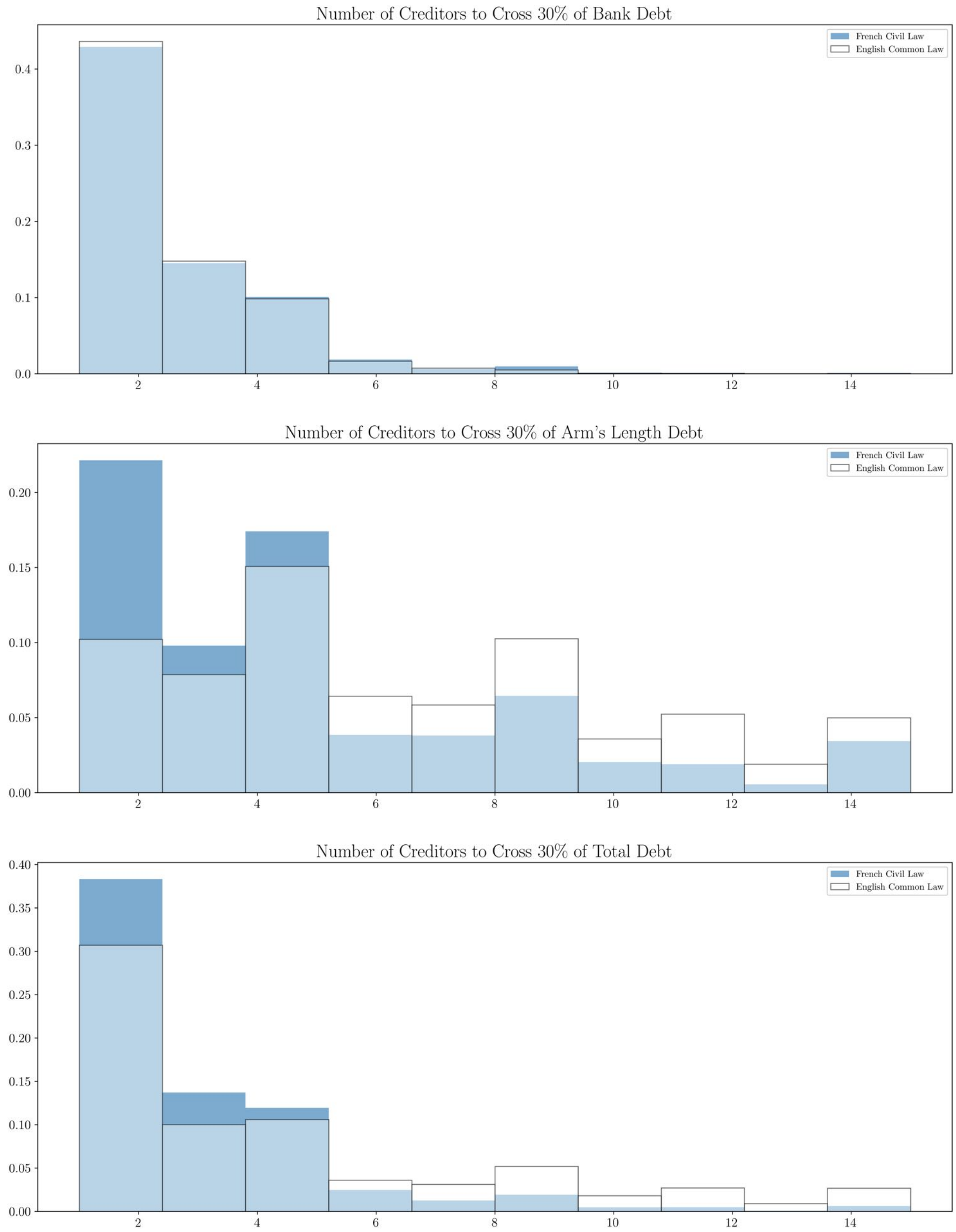


Figure 2.6: This figure displays the distribution of number of creditors to cross 30% of debt, separately for bank debt, arm's length debt, and total debt, comparing French Civil Law countries and English Common Law countries. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

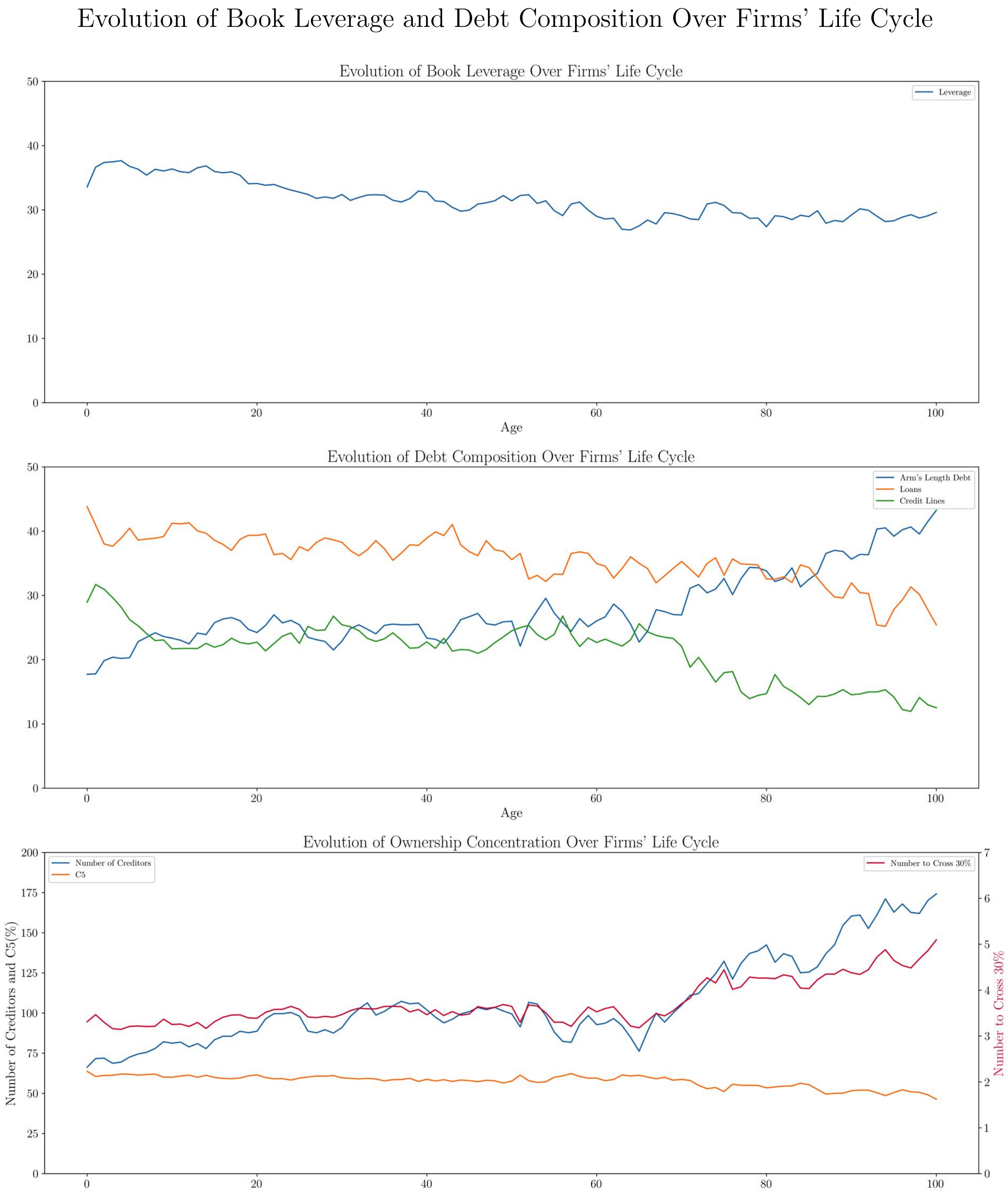


Figure 2.7: This figure displays the global evolution of book leverage, debt composition, and debt ownership concentration over firms' life cycle. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

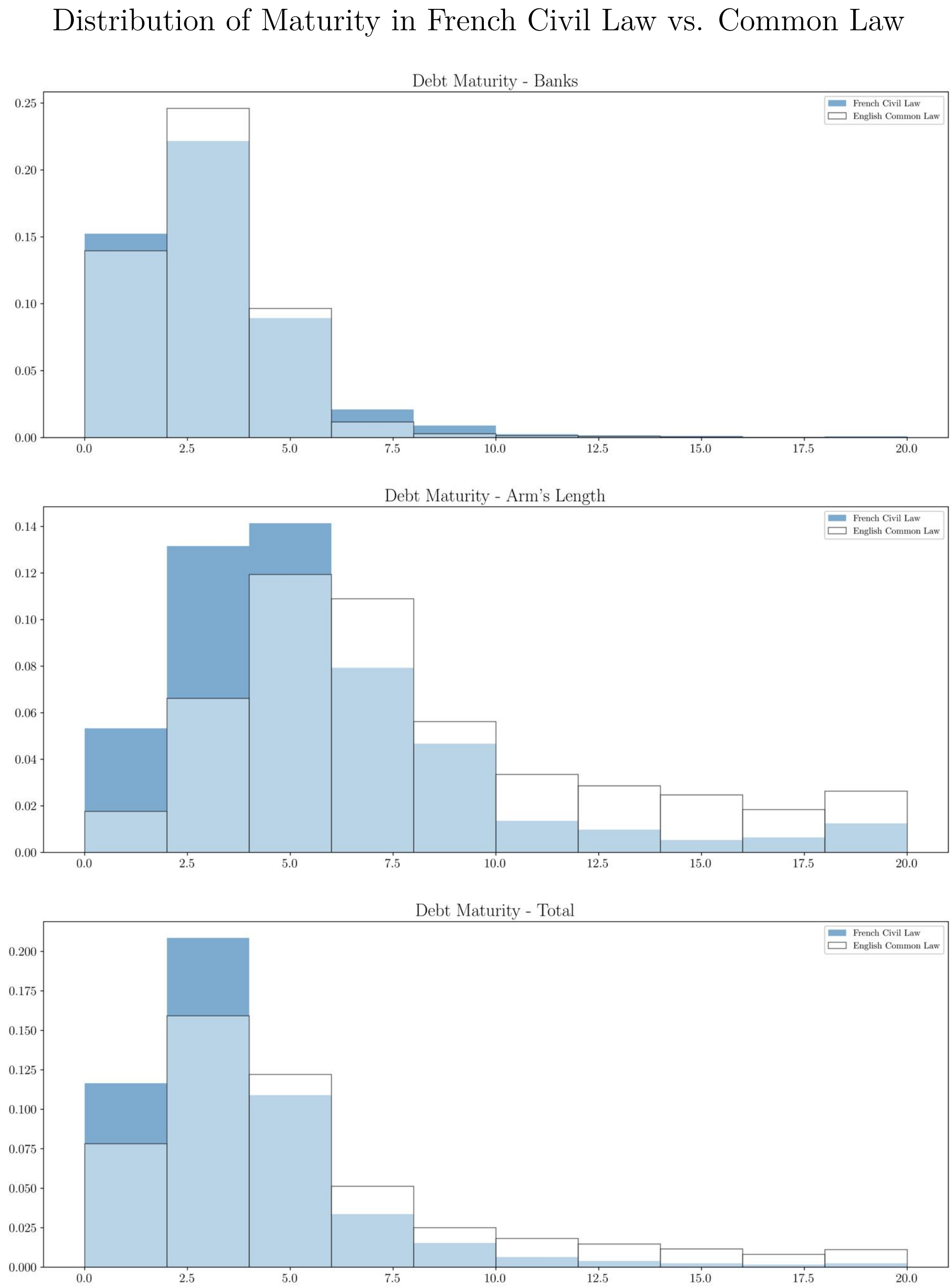


Figure 2.8: This figure displays the distribution of debt maturity, separately for bank debt, arm's length debt, and total debt, comparing French Civil Law countries and English Common Law countries. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

Distribution of Percentage of Secured Debt in French Civil Law vs. Common Law

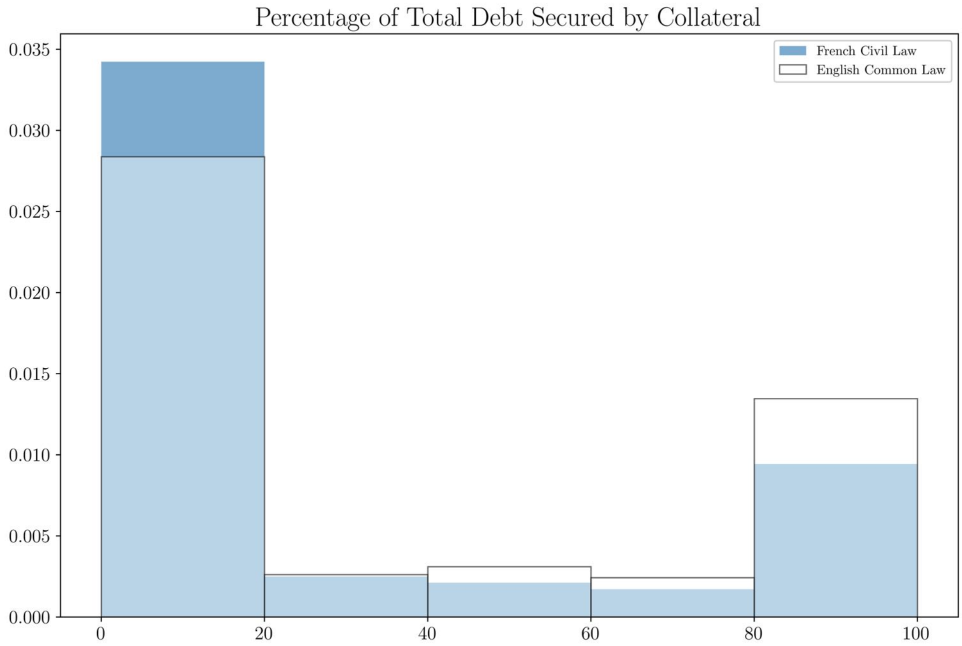


Figure 2.9: This figure displays the distribution of percentage of total debt that is secured by some collateral, comparing French Civil Law countries and English Common Law countries. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

Distribution of Debt Currency in French Civil Law vs. Common Law

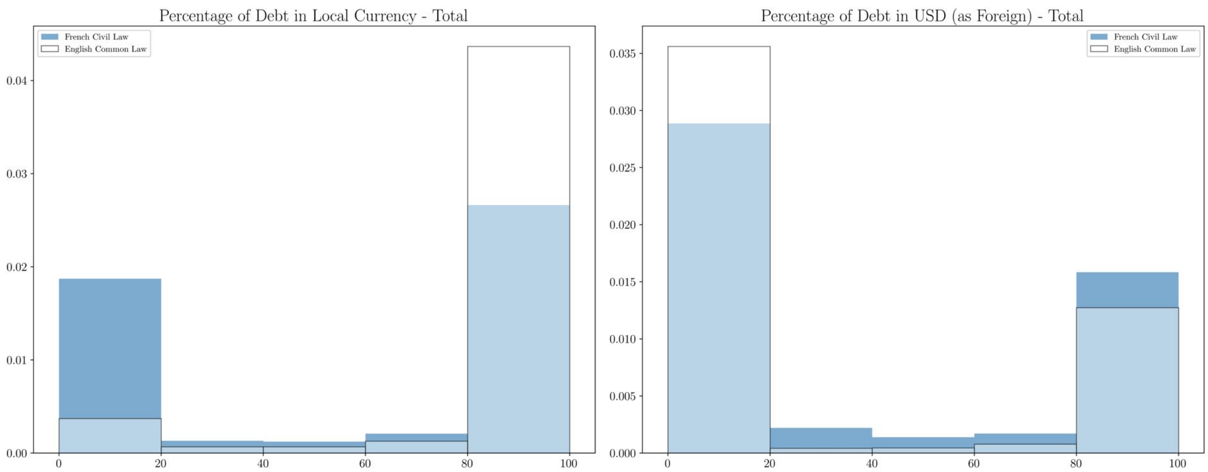


Figure 2.10: This figure displays the distribution of debt currency, comparing French Civil Law countries and English Common Law countries. Detailed variable definitions and data sources are provided in Appendix Section A. Averages are calculated over all years in the sample, and missing data are disregarded in calculating the averages.

Appendix B

Appendices

B.1 Appendix

A Variable Definitions

Table A1: Variable Description and Sources

Variable	Description	Source
Total Debt		
Arm's length debt	The variable represents the total outstanding amount of senior bonds and notes, subordinated bonds and notes, and commercial papers, expressed as a percentage of the total debt. This calculation is based on the identified debt instruments in eMAXX. Additionally, we have access to a total sum provided by Capital IQ. When mentioning the total sum from Capital IQ, we explicitly refer to it as CIQ_Arm's length debt.	eMAXX, Capital IQ, Worldscope, Compustat
Bank debt	The variable represents the total outstanding amount of term loans and drawn lines of credit, expressed as a percentage of the total debt. This calculation is based on the identified debt instruments in Dealscan. Additionally, we have access to a total sum provided by Capital IQ. We also utilized the total sum from Capital IQ to deduce the proportion of total allocated credit drawn in each quarter. When mentioning the total sum from Capital IQ, we explicitly refer to it as CIQ_Bank debt.	Dealscan, Capital IQ, Worldscope, Compustat

Continued

Table A1 – Continued

Variable	Description	Source
Loans Observed	The variable represents the outstanding amount of term loans recorded in Dealscan for a quarter, expressed as a percentage of the total debt. It is calculated by subtracting the installments paid up to the respective quarter from the total loan amount issued. In cases where repayment schedules are missing in Dealscan, we assumed a homogeneous repayment rate for loans to align with the total amount of term loans recorded in Capital IQ. If the sum of term loans in Dealscan is less than the total in Capital IQ, we assumed a bullet payment at maturity for loans with missing repayment schedules.	Dealscan, Capital IQ, Worldscope, Compustat
Lines Observed	The variable represents the outstanding amount of drawn lines of credit recorded in Dealscan for a quarter, expressed as a percentage of the total debt. We assumed a homogeneous drawing rate for all outstanding lines of credit relative to the total drawn amount indicated in Capital IQ.	Dealscan, Capital IQ, Worldscope, Compustat
Identified Instruments	The variable represents the sum of the outstanding amount of arm's length debt in eMAXX and bank debt in Dealscan, expressed as a percentage of the total debt.	eMAXX, Dealscan, Capital IQ, Worldscope, Compustat
Total Debt Ownership Observed	The variable represents the sum of the observed debt ownership of arm's length debt in eMAXX and bank debt in Dealscan, expressed as a percentage of the total debt.	eMAXX, Dealscan, Capital IQ, Worldscope, Compustat

Continued

Table A1 – Continued

Variable	Description	Source
Num Creditors	The variable represents the count of owners of arm's length debt in eMAXX and lenders in Dealscan. It's worth noting that the lenders in Dealscan are assumed to be different from the eMAXX bondholders, as banks typically do not appear in eMAXX. Additionally, we have excluded a few securitized tranches of loans from Dealscan that were also present in eMAXX.	eMAXX, Dealscan
Herfindahl of Creditors Ownership	The variable represents the sum of squared fractional holdings of total creditors. These fractions are calculated w.r.t. the total ownerships observed for the respective firm in a quarter. This variable are defined only if the observed ownerships accounts for at least 30% of total debt.	eMAXX, Dealscan
C1, C3, C5, C10	The variable represents the ownership stake of the largest k creditors, where k is equal to 1, 3, 5, and 10, regardless of the type of debt instrument they hold. This stake is expressed as a percentage of the total ownerships observed for the respective firm in a quarter. These variables are defined only if the observed ownerships accounts for at least 30% of total debt.	eMAXX, Dealscan
Num to cross 10%, 20%, 30%	The variable represents the number of creditors needed to surpass the threshold of k% of the total ownerships observed, where k is equal to 10, 20, and 30. This calculation is irrespective of the type of debt instrument held by these creditors. These variables are defined only if the observed ownerships accounts for at least 30% of total debt.	eMAXX, Dealscan

Continued

Table A1 – Continued

Variable	Description	Source
Largest Creditor Type	The variable represents the institutional creditor type that appeared most frequently as the largest creditor of firm-quarter observations within a country.	eMAXX, Dealscan
Total Maturity of debt	The variable represents the weighted average maturity of all identified debt instruments, regardless of their type. Maturities are measured in years (365 days), and weights are determined by the USD equivalent of the outstanding amount of debt instruments.	eMAXX, Dealscan
Total Secured Debt %	The variable represents the percentage of total debt mentioned as secured in Capital IQ.	Capital IQ
Local Currency %	The variable represents the percentage of total debt issued in the local currency of the firm's registered address.	Capital IQ
USD, EUR, GBP, JPY, CAD, AUD, NZD , NOK, SEK, CHF, Other Currencies %	The variable represents the percentage of total debt issued in USD, EUR, GBP, JPY, CAD, AUD, NZD, NOK, SEK, CHF, or other currencies.	Capital IQ, Worldscope, Compustat
Arm's Length Debt		
Bonds Avg size USD m	The variable represents the average size of the outstanding amount of arm's length debt instruments with a positive amount outstanding in a quarter, measured in USD millions.	eMAXX

Continued

Table A1 – Continued

Variable	Description	Source
Bonds Outstanding PAR % of Total Debt	The variable represents the outstanding amount of arm's length debt as a percentage of the total debt. This is equivalent to the variable arm's length debt defined earlier.	eMAXX, Worldscope, Compustat
Bond Ownership Observed % of Outstanding PAR	The variable represents the percentage of the total outstanding amount of bonds in a quarter that we observe ownership of in our sample.	eMAXX
Num of Bond Issues	The variable represents the number of arm's length debt instruments with a positive amount outstanding in a quarter.	eMAXX
Num of Bondholders	The variable represents the number of distinct owners of arm's length debt instruments of a firm in a quarter, regardless of their stake size.	eMAXX
Herfindahl of Bondholders Ownership	The variable represents the sum of squared fractional holdings of total owners of arm's length debt. These fractions are calculated w.r.t. the total ownerships observed of all identified arm's length debt instruments in eMAXX. This variables are defined only if the observed ownerships accounts for at least 30% of total outstanding amount of arm's length debt.	eMAXX
C1, C3, C5, C10	The variable represents the ownership stake of the largest k owners of arm's length debt, where k is equal to 1, 3, 5, and 10, regardless of the instrument they hold. This stake is expressed as a percentage of the total ownerships observed of all identified arm's length debt instruments in the respective quarter. These variables are defined only if the observed ownerships accounts for at least 30% of total outstanding amount of arm's length debt.	eMAXX

Continued

Table A1 – Continued

Variable	**Description**	**Source**
Num to cross 10%, 20%, 30%	The variable represents the number of arm's length creditors needed to surpass the threshold of k% of the total ownerships observed, where k is equal to 10, 20, and 30. This calculation is irrespective of the type of debt instrument held by these creditors. These variables are defined only if the observed ownerships accounts for at least 30% of total outstanding amount of arm's length debt.	eMAXX
Largest Bondholder Type	The variable represents the institutional type of arm's length creditors that appeared most frequently as the largest arm's length creditor of firm-quarter observations within a country.	eMAXX
Maturity of arm's length debt	The variable represents the weighted average maturity of all identified arm's length debt instruments. Maturities are measured in years (365 days), and weights are determined by the USD equivalent of the outstanding amount of arm's length debt instruments.	eMAXX
Arm's length Secured Debt %	The variable represents the percentage of total arm's length debt mentioned as secured in Capital IQ.	Capital IQ
Bank Debt		
Num of Tranches (loans & lines)	The variable represents the number of term loans and lines of credit with a positive amount outstanding in a quarter.	Dealscan
Num of Loans	The variable represents the number of term loans with a positive amount outstanding in a quarter.	Dealscan

Continued

Table A1 – Continued

Variable	Description	Source
Num of Lines of Credit	The variable represents the number of lines of credit with a positive amount outstanding in a quarter.	Dealscan
Num Lenders	The variable represents the number of all distinct lenders that hold a positive fraction of Dealscan tranches with a positive amount outstanding in a quarter.	Dealscan
Num Lead Arrangers	The variable represents the number of all distinct lead arrangers of Dealscan tranches with a positive amount outstanding in a quarter.	Dealscan
Herfindahl of Lenders Amounts	The variable represents the sum of squared fractional holdings of total owners of bank debt. These fractions are calculated w.r.t. the total outstanding amount of all identified bank debt instruments in Dealscan.	Dealscan
C1, C3, C5, C10	The variable represents the ownership stake of the largest k lenders of bank debt, where k is equal to 1, 3, 5, and 10, regardless of the instrument they hold. This stake is expressed as a percentage of the total outstanding amount of all identified bank debt instruments in the respective quarter.	Dealscan
Num to cross 10%, 20%, 30%	The variable represents the number of bank lenders needed to surpass the threshold of k% of the total bank ownerships observed, where k is equal to 10, 20, and 30. This calculation is irrespective of the tranche type held by these lenders.	Dealscan
Largest Lender Type	The variable represents the institutional type of bank creditors that appeared most frequently as the largest bank creditor of firm-quarter observations within a country.	Dealscan

Continued

Table A1 – Continued

Variable	Description	Source
Maturity of bank debt	The variable represents the weighted average maturity of all identified bank debt instruments. Maturities are measured in years (365 days), and weights are determined by the USD equivalent of the outstanding amount of all bank debt instruments.	Dealscan
Bank Secured Debt %	The variable represents the percentage of total bank debt mentioned as secured in Capital IQ.	Capital IQ
Country Measures		
Ex Ante Control Index	The ex ante private control of self-dealing index, ranging from 0 to 1, encompasses (i) disclosures by the buyer and seller, and (ii) requirements for a positive independent review of the transaction and approval by disinterested shareholders.	Djankov et al. (2008b)
Ex Post Control Index	The ex post private control of self-dealing index, ranging from 0 to 1, captures (i) post-transaction legal provisions for holding the buyer and seller liable for bad faith, (ii) shareholders' ability to sue or rescind the transaction, (iii) shareholders' ability to access evidence on the transaction, and (iv) disclosure of evidence in periodic filings.	Djankov et al. (2008b)
Anti self dealing Index	The variable represents the average of the ex ante and ex post control indexes, which ranges from 0 to 1.	Djankov et al. (2008b)
Public Enforcement Index	The variable represents the disclosure requirements for self-dealing transactions by managers and controlling shareholders, ranging from 0 to 1.	Djankov et al. (2008b)

Continued

Table A1 – Continued

Variable	**Description**	**Source**
Creditor Rights Index	The creditor rights index ranges from 0 to 4. A score of 1 is assigned for each of the following secured lenders' rights defined in laws and regulations: (i) restrictions, such as creditor consent or minimum dividends, for a debtor to file for reorganization, (ii) the ability of secured creditors to seize their collateral after the reorganization petition is approved (i.e., no automatic stay or asset freeze), (iii) the priority of secured creditors in being paid first out of the proceeds of liquidating a bankrupt firm, and (iv) management's loss of administration of the property pending the resolution of the reorganization. This index is derived from Djankov et al. (2007), who extend, revise, and update the original index compiled by La Porta et al. (1997, 1998).	Djankov et al. (2007)

B Appendix Tables

Table A2: Country Level Measures

Country	Creditor Rights	Ex Ante Control Index	Ex Post Control Index	Anti Self-dealing Index	Public Enforcement Index	Bankruptcy Efficiency Index
Australia	3.00	0.89	0.62	0.76	0.50	87.80
Bermuda	.	.	.	.	.	.
Canada	1.00	0.33	0.95	0.64	1.00	93.20
Cayman Islands	.	.	.	.	.	.
Guernsey	.	.	.	.	.	.
Hong Kong	4.00	1.00	0.93	0.96	0.00	88.30
India	2.00	0.33	0.82	0.58	0.50	.
Ireland	1.00	0.78	0.80	0.79	0.00	89.90
Israel	3.00	0.50	0.95	0.72	1.00	66.20
Jersey	.	.	.	.	.	.
Malaysia	3.00	1.00	0.90	0.95	1.00	48.40
New Zealand	4.00	1.00	0.90	0.95	0.00	90.70
Pakistan	1.00	0.17	0.65	0.41	0.75	.
Qatar	.	.	.	.	.	.
Saudi Arabia	3.00	.	.	.	.	40.60
Singapore	3.00	1.00	1.00	1.00	1.00	96.10
South Africa	3.00	1.00	0.62	0.81	0.00	39.80
Thailand	2.00	1.00	0.62	0.81	0.00	54.90
United Arab Emirates	2.00	.	.	.	.	21.80
United Kingdom	4.00	1.00	0.90	0.95	0.00	92.30
United States	1.00	0.33	0.97	0.65	0.00	85.80
English Common Law	**2.50**	**0.74**	**0.83**	**0.79**	**0.41**	**71.13**
Denmark	3.00	0.25	0.68	0.46	0.75	76.70
Finland	1.00	0.14	0.78	0.46	0.00	92.40
Norway	2.00	0.42	0.42	0.42	1.00	91.80
Sweden	1.00	0.17	0.50	0.33	1.00	86.00
Scandinavian Civil Law	**1.75**	**0.24**	**0.59**	**0.42**	**0.69**	**86.73**
Austria	3.00	0.00	0.42	0.21	1.00	78.00
China	2.00	1.00	0.53	0.76	0.00	43.60
Germany	3.00	0.14	0.42	0.28	1.00	57.00
Japan	2.00	0.22	0.78	0.50	0.00	95.50
Poland	1.00	0.25	0.33	0.29	1.00	67.70
South Korea	3.00	0.25	0.69	0.47	0.50	88.10
Switzerland	1.00	0.08	0.45	0.27	0.75	60.40
Taiwan	2.00	0.42	0.71	0.56	0.00	93.80
German Civil Law	**2.12**	**0.30**	**0.54**	**0.42**	**0.53**	**73.01**
Argentina	1.00	0.33	0.35	0.34	0.00	35.80
Belgium	2.00	0.39	0.70	0.54	0.50	90.80
Brazil	1.00	0.22	0.33	0.27	0.50	13.40
Chile	2.00	0.50	0.75	0.62	1.00	40.90
Colombia	0.00	0.83	0.31	0.57	0.00	64.80
France	0.00	0.08	0.68	0.38	0.50	54.10
Greece	1.00	0.08	0.35	0.22	0.50	53.80
Indonesia	2.00	0.81	0.50	0.65	0.00	25.10
Italy	2.00	0.17	0.68	0.42	0.00	45.30
Kuwait	3.00	.	.	.	.	55.90
Luxembourg	.	0.17	0.40	0.28	1.00	.
Mexico	0.00	0.19	0.15	0.17	0.50	72.60
Netherlands	3.00	0.06	0.35	0.20	0.00	94.90
Oman	0.00	.	.	.	.	53.50
Peru	0.00	0.25	0.65	0.45	0.25	41.80
Philippines	1.00	0.06	0.38	0.22	0.00	17.50
Russia	2.00	0.81	0.07	0.44	1.00	39.00
Spain	2.00	0.22	0.53	0.37	1.00	82.00
Turkey	2.00	0.33	0.53	0.43	0.00	6.60
French Civil Law	**1.33**	**0.32**	**0.45**	**0.39**	**0.40**	**49.32**
World	**1.91**	**0.45**	**0.61**	**0.53**	**0.45**	**63.97**

This table presents the country level measures of legal environment if different countries. Please discuss Table A for definitions.

Table A3: Creditor Concentration- C10

	(1) C10 Creditors	(2) C10 Arm's	(3) C10 Bank	(4) C10 Creditors	(5) C10 Arm's	(6) C10 Bank	(7) C10 Creditors	(8) C10 Arm's	(9) C10 Bank
French	7.40***	4.86*	-0.70						
	(2.35)	(2.73)	(0.95)						
German	8.06**	22.82**	-1.04						
	(3.91)	(9.76)	(2.18)						
Scandinavian	12.87***	10.92***	2.68***						
	(1.84)	(2.45)	(0.37)						
anti self dealing Index				-8.36	-8.16	2.08			
				(10.29)	(7.33)	(2.20)			
SizeGroup=0 × anti self dealing Index							-43.35***	-3.53	-11.35***
							(9.86)	(14.17)	(4.14)
SizeGroup=1 × anti self dealing Index							-12.79	2.10	-0.23
							(9.72)	(8.92)	(3.30)
SizeGroup=2 × anti self dealing Index							-6.41	-4.06	2.43
							(9.13)	(8.18)	(2.47)
SizeGroup=3 × anti self dealing Index							-9.87	-8.71	1.60
							(10.88)	(7.72)	(2.31)
log assets	-11.20***	-10.61***	-4.77***	-11.58***	-10.72***	-4.71***	-12.52***	-9.90***	-5.23***
	(1.53)	(0.19)	(0.45)	(1.31)	(0.11)	(0.42)	(0.84)	(0.37)	(0.40)
log age	-1.25***	0.00	0.43**	-0.76*	0.11	0.42**	-0.63	0.16	0.46**
	(0.30)	(0.23)	(0.19)	(0.45)	(0.17)	(0.18)	(0.44)	(0.18)	(0.18)
Log GDP per capita	-1.90*	-3.62	0.79	-3.53***	-6.38***	1.07**	-3.16**	-6.76***	1.27**
	(0.97)	(2.19)	(0.48)	(1.23)	(1.54)	(0.51)	(1.55)	(1.67)	(0.48)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All
Adj. R2	0.571	0.570	0.250	0.563	0.557	0.248	0.579	0.560	0.256
N	122049	46841	101452	119896	46474	99494	119896	46474	99494

The table investigates the association of ownership dispersion, measured by the ownership stake of largest ten creditors, in their debt type (all, arm's length debt, or bank debt) with the legal environment. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A4: Creditor Concentration- HHI

	(1) HHI Total	(2) HHI Arm's	(3) HHI Bank	(4) HHI Total	(5) HHI Arm's	(6) HHI Bank	(7) HHI Total	(8) HHI Arm's	(9) HHI Bank
French	-0.00	0.01	-0.02						
	(0.01)	(0.02)	(0.01)						
German	-0.04	0.23**	-0.07*						
	(0.04)	(0.10)	(0.04)						
Scandinavian	0.04*	0.01	0.05**						
	(0.02)	(0.01)	(0.02)						
anti self dealing Index				0.04	-0.02	0.05			
				(0.04)	(0.04)	(0.04)			
SizeGroup=0 × anti self dealing Index							0.41**	0.48**	0.43**
							(0.20)	(0.19)	(0.17)
SizeGroup=1 × anti self dealing Index							0.20**	0.15***	0.20***
							(0.08)	(0.05)	(0.07)
SizeGroup=2 × anti self dealing Index							0.05	-0.01	0.06
							(0.06)	(0.05)	(0.05)
SizeGroup=3 × anti self dealing Index							0.03	-0.02	0.04
							(0.05)	(0.04)	(0.04)
log assets	-0.08***	-0.03***	-0.07***	-0.08***	-0.03***	-0.07***	-0.05***	-0.02***	-0.04***
	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)
log age	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Log GDP per capita	-0.00	-0.02	0.01	0.00	-0.04***	0.02	-0.01	-0.04***	0.01
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All
Adj. R2	0.325	0.245	0.217	0.324	0.189	0.205	0.351	0.215	0.231
N	122049	46841	101452	119896	46474	99494	119896	46474	99494

The table investigates the association of ownership dispersion, measured by Herfindahl–Hirschman index in their debt type (all, arm's length debt, or bank debt) with the legal environment. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

Table A5: Leverage and Debt Composition 1

	(1) Leverage	(2) Arm's length Debt % Total Assets	(3) Bank Debt % Total Assets	(4) Leverage	(5) Arm's length Debt % Total Assets	(6) Bank Debt % Total Assets	(7) Leverage	(8) Arm's length Debt % Total Assets	(9) Bank Debt % Total Assets
French	-0.00	-5.02***	4.62**						
	(0.02)	(1.65)	(1.80)						
German	-0.04*	-6.69***	3.26***						
	(0.02)	(1.28)	(0.98)						
Scandinavian	0.02	-4.77***	6.46***						
	(0.02)	(1.68)	(0.90)						
anti self dealing Index				-0.04	4.34	-8.08**			
				(0.05)	(6.81)	(3.35)			
SizeGroup=0 × anti self dealing Index							-0.08	10.58	-17.21***
							(0.07)	(6.46)	(3.29)
SizeGroup=1 × anti self dealing Index							-0.08	3.28	-10.48***
							(0.05)	(6.39)	(3.08)
SizeGroup=2 × anti self dealing Index							-0.04	4.50	-7.41**
							(0.05)	(6.50)	(2.92)
SizeGroup=3 × anti self dealing Index							-0.05	4.41	-8.50**
							(0.05)	(7.15)	(3.45)
log assets	0.00	3.13***	-3.26***	0.00*	3.44***	-3.42***	-0.00	3.53***	-3.73***
	(0.00)	(0.50)	(0.44)	(0.00)	(0.35)	(0.36)	(0.00)	(0.29)	(0.34)
log age	-0.02***	0.62***	-2.33***	-0.02***	0.22	-2.04***	-0.02***	0.20	-2.01***
	(0.00)	(0.20)	(0.45)	(0.00)	(0.25)	(0.30)	(0.01)	(0.25)	(0.31)
Log GDP per capita	-0.00	2.12***	-2.28**	0.01	3.50***	-2.94***	0.01	3.46***	-2.80***
	(0.01)	(0.63)	(1.05)	(0.01)	(1.05)	(0.78)	(0.01)	(1.28)	(0.82)
Fixed effects	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q	I & Q
Cluster	C	C	C	C	C	C	C	C	C
Sample	All	All	All	All	All	All	All	All	All
Adj. R2	0.160	0.285	0.263	0.152	0.269	0.258	0.154	0.271	0.261
N	122049	122049	122049	119896	119896	119896	119896	119896	119896

The table investigates the association of broad capital structure variables with the legal environment. Leverage is defined as the ratio of total debt to total assets. Other variables have clear definition in the column headers. The key explanatory variables include legal origin indicator variables, which take the value of 1 for French civil-law, German civil-law, and Scandinavian civil-law countries, respectively, with common-law legal origin serving as the omitted category, in addition to the composite anti-self-dealing index, which represents the average of the ex ante and ex post private control of self-dealing measures according to Djankov et al. (2008b). This index ranges from 0 to 1; SizeGroup includes the quartile of firms' size (total assets in USD) in the population of firms in that country. All specifications incorporate industry fixed effects using two-digit SIC classifications and year × quarter fixed effects. Controls for the logarithm of GDP per capita for a given year in each country, the logarithm of firm age and the logarithm of total assets in USD are included in all specifications. The constant term is not reported. Detailed variable definitions and sources can be found in Appendix Section A. Heteroskedasticity-adjusted standard errors, clustered at the country level, are reported in parentheses below the estimates. Symbols ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

C Data preparation

Firm Identifiers

Firm information within our sample is sourced from four primary databases: Capital IQ, Worldscope, Compustat, and Amadeus. These databases utilize different firm identifiers: companyid for Capital IQ, Worldscope Permanent Identifier for Worldscope, GVKEY for Compustat, and IDNR for Amadeus. Of these, Capital IQ stands out as the most comprehensive dataset, and its companyid serves as our primary firm identifier. Therefore, we try to find links between companyid and the local identifiers of the other three datasets. Below, we outline the steps taken to merge each of these datasets with Capital IQ:

Capital IQ. Among the three sources of balance sheet information in our dataset, merging companyid with Compustat proves the simplest, facilitated by both being part of S&P Global, with a linking table provided directly by Capital IQ. The merge between Capital IQ and Amadeus, while challenging in practice, is straightforward to describe due to its predominantly manual hand-matching nature. Detailed procedures for these two mergers of datasets are elaborated below. However, the merge between Worldscope (from Refinitiv environment) and Capital IQ (from S&P Global) presented notable complexities. Despite both being reputable data providers, their distinct environments necessitated a more intricate merging process. The problem arises from Worldscope permanent identifier not appearing anywhere else than Worldscope. There is also not much hope in merging on public identifiers such ISIN and CUSIP directly because many firms in our sample either don't have them or don't have a unique value for them. As a result, we have to find global linking tables between S&P Global and Refinitiv environments.

The unique identifier common across all datasets within Refinitiv is known as PermID[1][2]. PermID serves as a comprehensive identifier for various entity types, including organizations, instruments, funds, issuers, and individuals. It remains unchanged and ensures unambiguous referencing, making it highly suitable as a universal identifier. PermID plays a central role in Refinitiv's information model and knowledge graph[3].

Our primary use of PermID is to merge Capital IQ with Worldscope, as detailed below. Additionally, having PermID facilitates the merging of firm data with debt ownership data in Dealscan (which now includes PermID in its revamped version) and eMAXX (for which PermIDs are retrieved similarly to Worldscope using Refinitiv Python APIs) which are both products of Refinitiv, as outlined in Appendix Section C.

There is no direct linkage between companyids in Capital IQ and PermIDs in the Refinitiv environment. To bridge this gap, we extensively utilize local linking tables provided in WRDS for each dataset. These tables detail the identifiers of instruments (such as bonds and stocks) associated with each firm. Specifically, we leverage the following linking tables from Capital IQ[4]:

1. *wrds_isin.sas7bdat*: Lists all ISINs associated with a companyid, including start and end dates of the linkage, and company names.

2. *wrds_cusip.sas7bdat*: Provides all CUSIPs and CINSs associated with a companyid, along with start and end dates of the linkage, and company names.

In contrast, for Refinitv PermID, a direct linking table from standard identifiers (ISIN, CUSIP, etc.) to organization PermID does not exist and must be compiled using available Refinitiv files[5]:

[1] Although sometimes PermID is not readily available in some Refinitiv datasets, as is the case with Worldscope, and needs to be retrieved separately using Refinitiv APIs or desktop products.

[2] See all details in https://permid.org/

[3] LSEG completed the acquisition of Refinitiv from its previous owners in late January 2021, and Refinitiv now operates as a subsidiary of LSEG. While some products of Refinitiv have undergone name changes since our original use, we continue to refer to them as Refinitiv to maintain clarity.

[4] These linking tables are located in the following directories within the SAS version of WRDS files: /wrds/capitaliq/sasdata/helper.

[5] These files are found in the following directories within the SAS version of WRDS files: /wrds/tfn/sasdata/helper

1. *permisindata.sas7bdat*: Shows the PermID of the instrument associated with each ISIN, detailing the start and end date of the link.

2. *perminstrref.sas7bdat*: Links each instrument PermID to its organization's PermID.

Similar steps are repeated using *permcusipdata.sas7bdat* and *permcincusipdata.sas7bdat* to link CUSIPs and CINSs to firm PermID. This results in two tables linking standard identifiers (ISIN, CUSIP, CINS) to their respective firm PermID, akin to the tables found in Capital IQ.

In the final step, we merge the companyid-ISIN linking table with the PermID-ISIN linking table[6] to establish companyid-PermID links. This process is similarly applied to companyid-CUSIP/CINS links and PermID-CUSIP/CINS links to refine our initial companyid-PermID associations. Finally, cases where the companyid-PermID relationship is not one-to-one are filtered out.

Following this procedure, out of 135,807 unique companyids in our Capital IQ dataset, we successfully identify PermIDs for 107,545 firms. This success rate is significant, with unmatched firms likely being smaller entities not covered by Refinitv PermIDs, which would not be included in our final sample anyway as our ownership data exclusively relies on Refinitiv products.

Worldscope. To compile a comprehensive dataset from Worldscope, we initially download all available observations covering the period 1998 to 2022, filtering out entries lacking a Worldscope Permanent Identifier or Date of Fiscal Period End. Subsequently, we exclude observations where the Entity Type is not designated as "C", indicating non-corporate entities, and remove entries with non-positive total assets. This refining process results in approximately 2.5 million quarterly observations.

[6]**Note:** At this stage, we are agnostic to the security types of ISINs or their existence in our debt instruments, aiming to maximize the number of links between Capital IQ and Refinitiv firm identifiers. Similarly, this approach applies to CUSIPs and CINS. Therefore, links may be established using, for example, stocks, even though our study focuses exclusively on debt instruments, primarily bonds and loans.

The Worldscope Permanent Identifier, unique to Worldscope and not shared with other datasets, precludes its direct use for merging. However, leveraging Refinitiv's ecosystem, we utilize their APIs to retrieve PermID for Worldscope firms. Initially, we employ the *Eikon get_symbology* method, focusing on Worldscope Permanent Identifiers. For firms where PermID retrieval is unsuccessful in the first step, we extend the query to utilize ISIN, CUSIP, and SEDOL identifiers in Worldscope.

Combining these approaches ensures we obtain PermIDs for approximately 99.8% of Worldscope observations. The remaining small fraction, where PermIDs cannot be retrieved confidently, is subsequently omitted from our dataset. Ultimately the merge of Capital IQ with Worldscope is done using PermIDs and calendar quarters.

Compustat. Compustat and Capital IQ both belong to S&P Global, and a linking table between Capital IQ companyids and Compustat GVKEYs is directly provided by Capital IQ, which is also accessible in WRDS[7]. This table includes the start date and end date of the validity of these links as well.

To obtain the GVKEYs for our quarterly Capital IQ firm observations, we proceed with the following steps:

1. First, we merge our Capital IQ data with the linking table, using only companyids that are mapped to a unique GVKEY.

2. For companyids mapped to more than one GVKEY, we perform the merge again, this time taking into account the validity periods of the links.

Ultimately the merge between Capital IQ and Compustat is performed on the GVKEYs and calendar quarters.

[7]This linking table is called *wrds_gvkey.sas7bdat* and is located in the following directories within the SAS version of WRDS files: /wrds/capitaliq/sasdata/helper

Amadeus. Amadeus provides balance sheet information on European firms using a proprietary local identifier called IDNR, which is not compatible with other datasets. The only standard identifier in Amadeus is ISIN, which only sporadically adds firms to our sample that are not present already in Worldscope or Compustat. To enhance our sample coverage in Europe, we employ the following approach:

1. We utilize our Dealscan data, which includes PermID and GVKEYs[8] of borrowers, to identify European firms:

    - Whose PermID is not in Worldscope,
    - Either lack a GVKEY or have a GVKEY not in our Compustat data,
    - Whose PermID appears in our Capital IQ data.

    On one hand, these firms would not appear in our final dataset without using Amadeus, as their balance sheet information cannot be sourced from either Worldscope or Compustat. On the other hand, these firms have highest potential for inclusion in our study because their debt structure already exists in our Capital IQ data, and some debt ownership data for them exists in our Dealscan sample. We identify 1,063 such Dealscan borrowers.

2. We manually search for firm names online and in Amadeus and locate 572 out of the 1,063 firms in Amadeus.

3. For these 572 located firms in Amadeus, we fill in their PermID and merge Capital IQ with Amadeus using ISIN and then PermID. Although in general we only keep clean Capital IQ observation with detail breakdown of debt types, for this 572 firms in Amadeus we tolerate missing detail debt structure information in Capital IQ and we assume they are fully bank financed. However, exclusion of these 572 firms from

[8]GVKEY of Dealscan borrowers are found using the table provided by Chava and Roberts (2008)

our sample does not qualitatively or quantitatively affect the material findings of this study.

eMAXX

eMAXX data reports institutional investors' fixed-income holdings at the securities level. Based on the calendar date of bond ownership, this dataset is structured as separate standalone folders for each quarter. The unit of observation in eMAXX is fund, so we pivot the date to achieve a dual format in which we observe different holder of each security in each quarter.

We begin by compiling a list of issue-level and issuer-level data from all the eMAXX folders available to us, retaining only the records pertaining to the corporate and asset-backed markets. Upon analysis, we observed that many entries regarding the par amount outstanding reported in eMAXX for corporate issues were outdated, as numerous securities had been partially or fully called before reaching maturity. Unfortunately, these updates were either not reflected in eMAXX or were significantly delayed. Since the par outstanding of bonds is a crucial piece of information for reconstructing the debt outstanding of bond-issuing firms, we addressed this issue by retrieving the time-series data of par outstandings for all eMAXX issues using the Refinitiv API and subsequently updating the information in eMAXX. At this stage, we are dealing with approximately half a million distinct CUSIPs in our dataset.

The next step, perhaps the most challenging aspect of working with eMAXX data, involves linking eMAXX issuers to other datasets containing firm information, such as Capital IQ, Worldscope, Compustat, and Amadeus. The only identifiers available in eMAXX for issuers are the 6-digit CUSIP and the firm name, both of which can be cumbersome to work with. Dealing with 6-digit CUSIPs can be challenging because a firm may be assigned multiple 6-digit CUSIPs over time, and handling firm names can be complicated due to potential variations in recording across datasets and the presence of overlapping or similar names. To tackle this challenge, we employ a three step solution: Firstly, we utilize the auxiliary Capital IQ linking tables available in the SAS studio version of WRDS, which provide historical changes in CUSIP-companyid and ISIN-companyid links. Secondly, we utilize auxiliary Thomson-Reuters files, also available in the SAS studio version of WRDS, which give us the PermID of the securities in eMAXX. Although these are only the PermIDs of the bond issues, they help us find the PermID of the issuer using the Refinitiv API. Lastly, for bond issues that still lack either the Capital IQ companyid or Refinitiv PermID, we attempt to find additional firm PermIDs by searching the issue CUSIP, ISIN, or CINS in the Refinitiv API, which occasionally provides us with extra firm PermIDs. We consider the aggregate of these three steps very successful, as only about 11% of our half a million CUSIPs remain unlinked to both Capital IQ companyid and Refinitiv PermID.

Identifying and consolidating ownership information of bond issues is a crucial aspect of this research. However, this task can become challenging when ownership stakes are reported using seemingly distinct identifiers. This issue of identifier duplicity is particularly common when different segments of the same debt instrument are traded in both the US and Europe such as 144A and Regulation S bonds. Failing to establish these linkages could lead researchers to believe that eMAXX covers only a marginal portion of certain bonds' ownership, without realizing that another segment of bondholders for that same issue may be using an alternative identifier. To illustrate, consider a scenario where two distinct CUSIPs are associated with the same bond. In eMAXX holdings data, these identifiers might be observed with ownership percentages of x_A and x_B relative to the total par outstanding. In such cases, the correct percentage of observed ownership for this bond should be $x_A + x_B$. The reasons for the existence of such duplicate identifiers, the regulations surrounding them, and some examples are outlined in Section D. However, in the rest of this section, we elaborate on the methods for identifying and handling such duplicate issue identifiers.

To identify duplicating identifiers, we initially utilize the Refinitiv API to acquire the full list of identifiers for each bond issue in eMAXX. Among these identifiers, our focus is on CINS, ISIN, CUSIP, Corresponding 144A CUSIP, Corresponding 144A ISIN, Corresponding Regulation S CUSIP, and Corresponding Regulation S ISIN. For each security CUSIP in eMAXX, we meticulously compile a list of all potential other identifiers by examining not only the information provided for the respective CUSIP from the API, but also considering any other eMAXX CUSIPs that had this respective CUSIP as their potential duplicates.

After identifying the complete list of potential other identifiers, we set aside issues with only one unique identifier. For the remaining issues, we focus on bundles of two and three identifiers, as they are easier to resolve. Within these sets, we attempt to match a few simple cases. Many of these issues are privately placed bonds that are not initially eligible for trading by public American investors, at least for a certain period after the initial issue. A common scenario in this group is when a security is issued with separate 144A and Regulation S identifiers: the former for selling to American Qualified Institutional Buyers and the latter for selling to overseas (mostly European) investors. After a certain period, if the 144A security becomes qualified to be registered with the SEC, both the Registered and 144A versions may be entirely or partially exchanged for the Registered bond. To illustrate the issue, consider Table A6, which depicts the amount outstanding associated with a bond issued on September 30, 2009, by ACCO BRANDS CORP, with a total par outstanding of USD 460 million and a maturity date of March 15, 2015. This bond was partially sold under the Regulation S CUSIP (or CINS) U00445AB9 to overseas investors and partially sold under the 144A CUSIP 00081TAC2 to American Qualified Institutional Buyers (QIBs) and both of these issues were fully exchanged to CUSIP 00081TAD0 on May 14, 2010.

The issue that prompted us to identify and address the problems of duplicate bond issues is twofold: Firstly, if a researcher simply wants to determine the total par outstanding of bonds for ACCO BRANDS CORP in, for example 2009Q3, and they naively sum up all the par values associated with different bond CUSIPs of this firm, he may arrive at a total greater than the balance sheet's total amount of outstanding bonds, as he fails to recognize that two seemingly different CUSIPs 00081TAC2 and U00445AB9 actually belong to the same bond. Secondly, and perhaps more importantly for our research, many funds in eMAXX continue reporting their holdings using the identifier by which they initially purchased the security, even if a new CUSIP and ISIN are assigned to their bond. By addressing this issue, if we observe a bondholder claiming ownership of CUSIP 00081TAC2 or U00445AB9 in quarters after 2010Q2, we would understand that they hold CUSIP 00081TAD0 in reality.

Table A6: Example of Duplicate Identifiers - ACCO BRANDS CORP

quarter	00081TAD0	00081TAC2	U00445AB9
	Registered	144A	Regulation S
2009Q3	0	460	460
2009Q4	0	460	460
2010Q1	0	460	460
2010Q2	460	0	0
2010Q3	460	0	0
2010Q4	460	0	0
2011Q1	460	0	0
2011Q2	454.9	0	0
2011Q3	425.14	0	0
2011Q4	425.14	0	0
2012Q1	425.14	0	0
2012Q2	13.415	0	0
2012Q3	0	0	0

This table shows the amount outstanding associated with a bond issued on September 30, 2009, by ACCO BRANDS CORP, with a total par outstanding of USD 460 million and a maturity date of March 15, 2015 that had three associated Identifiers (CUSIPs.) The Identifiers and their type are shown for each bond along with their par outstanding in USD millions.

D Bond identifiers

Regulations

A private placement is a type of securities offering made to a small group of selected investors rather than the general public. In a private placement, the securities are typically sold to institutional investors such as banks, insurance companies, hedge funds, and other large investors. Private placements are typically exempt from the registration requirements of the Securities Act of 1933, which means that issuers can avoid the time, cost, and regulatory burden of a public offering. Instead, issuers can negotiate the terms of the offering directly with the investors, including the price, interest rate, maturity date, and other securities features.

In response to the rising internationalization of capital markets, the SEC issued Regulation S and Rule 144A in April 1990. Before 1990, few foreign corporations entered the US public debt markets because the SEC's registration and disclosure procedures were cumbersome. Meanwhile, the private debt market in the United States was essentially illiquid since investors had to keep assets for at least two years before reselling them. Regulation S and Rule 144A provided exceptions for this waiting time and SEC registrations, conditional on the buyers of these issues. The essential distinction between the two rules is found in their respective jurisdictions and geographical bounds. Regulation S oversees both domestic and international issuers' registration of offshore placements and offers a safe harbor for securities placement in non-US markets. It also permits these issues to be traded on various foreign markets. However, Rule 144A only applies to privately placed securities issued in the United States and sold to US Qualified Institutional Buyers (QIBs).

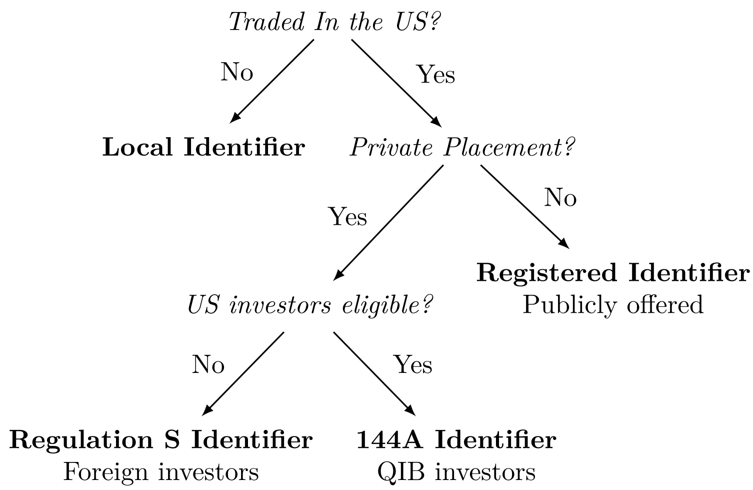


Figure A1: Bond Identifiers

Bond identifiers

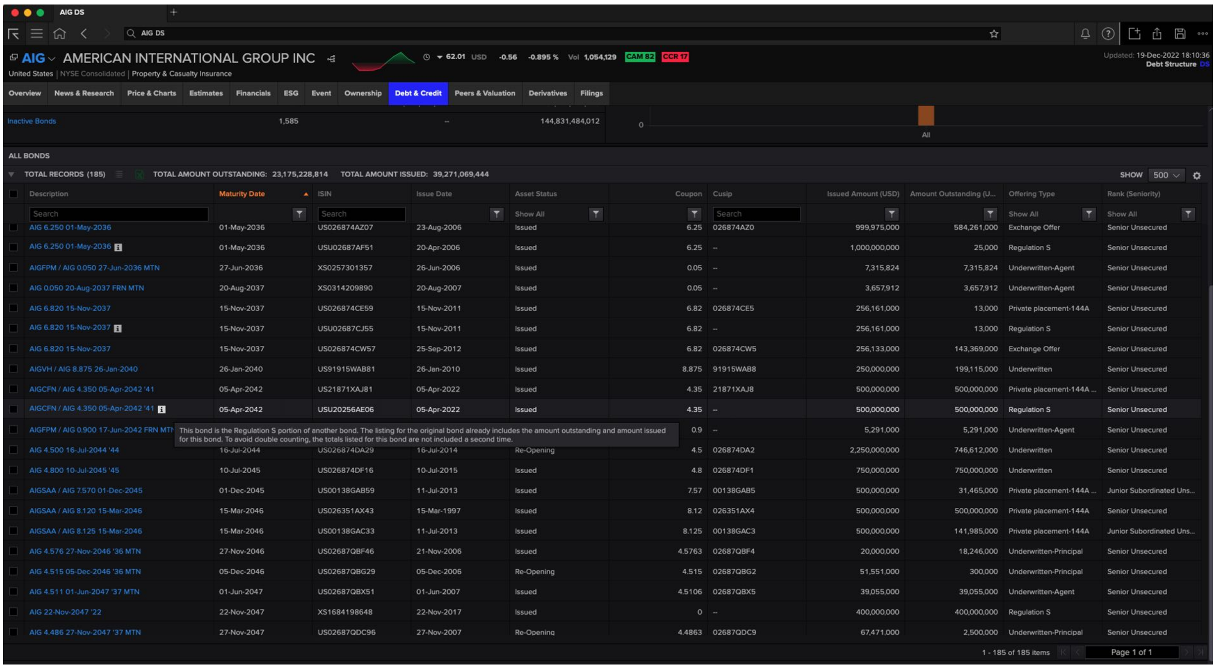
To correctly group multiple identifiers of the same bond issue, the first step is understanding why some securities have multiple identifiers. Two following cases may result in a multiplicity of identifiers. First, when the name CUSIP (the 6-digit CUSIP, also referred to as issuer's CUSIP) of the issuer changes, so does the CUSIP and ISIN of its issues. For example, if company A acquires company B at time t, the outstanding bond issues of B will be exchanged for an equivalent outstanding issue from A, with a new identifier pointing to A.

Second, depending on the eligible investor universe, the same bond issue might have multiple ISINs. Figure A1 shows the most common cases in which multiple active identifiers can point to the same issue. A bond not traded in the US will have one unique local set of identifiers. If it is traded in the US without private placement (equivalently if this bond is publicly traded), then it has to be registered with SEC, and it will get a registered CUSIP. However, suppose it is traded in the US only through private placement. In that case, the set of assigned identifiers depends on whether it will be sold to US or foreign investors.

If one issue only fits in one of the four subgroups in Figure A1, the observed identifier in eMAXX uniquely identifies the issue with certainty. The complication arises when one issue falls into two or more subgroups simultaneously. For example, a privately placed issue under rule 144A can later be registered with SEC and offered to the public. If it happens, this issue will have one 144A ISIN and one registered ISIN. Section D in Appendix lays down an example of such coupled securities, their accounting of their amount outstanding, and their identifiers for an example firm.

Regulation S securities can be offered in two ways: 1) as standalone Regulation S issues or 2) as concurrent Regulation S and Rule 144A offers, which differ in several respects. The investor base changes between the two segments: solo issues can only be offered to offshore investors, but combined issues can be sold to non-US offshore investors and US Qualified Institutional Buyers (QIBs). The main distinction is that the tranches of combined issues are subject to various jurisdictional limitations since offers outside the United States are subject to Regulation S, while those within the United States are subject to Rule 144A. The issuance and issuer features of the two types of issuances differ as well. Low credit grade issuers from the United States or nations geographically adjacent to the United States, such as Canada, Brazil, or Mexico, usually issue joint offerings. On the other hand, a typical solo issuance is often provided by a good credit quality issuer from a European country Li et al. (2021).

Crucially, security with both Regulation S and Rule 144A tranches is treated as one security in par value to prevent double-counting in databases and bond indices. In a similar manner, securities that originated under Rule 144A with registration rights and later registered with the SEC are treated as the same security for the treatment of the amount outstanding. We also address this issue by aggregating the bond ownership data in eMAXX.



Description	Maturity Date	ISIN	Issue Date	Asset Status	Coupon	Cusip	Issued Amount (USD)	Amount Outstanding (U...	Offering Type	Rank (Seniority)
AIG 6.250 01-May-2036	01-May-2036	US026874AZ07	23-Aug-2006	Issued	6.25	026874AZ0	999,975,000	584,261,000	Exchange Offer	Senior Unsecured
AIG 6.250 01-May-2036	01-May-2036	USU02687AF51	20-Apr-2006	Issued	6.25	--	1,000,000,000	25,000	Regulation S	Senior Unsecured
AIGFPM / AIG 0.050 27-Jun-2036 MTN	27-Jun-2036	XS0257301357	26-Jun-2006	Issued	0.05	--	7,315,824	7,315,824	Underwritten-Agent	Senior Unsecured
AIG 0.050 20-Aug-2037 FRN MTN	20-Aug-2037	XS0314209890	20-Aug-2007	Issued	0.05	--	3,657,912	3,657,912	Underwritten-Agent	Senior Unsecured
AIG 6.820 15-Nov-2037	15-Nov-2037	US026874CE59	15-Nov-2011	Issued	6.82	026874CE5	256,161,000	13,000	Private placement-144A	Senior Unsecured
AIG 6.820 15-Nov-2037	15-Nov-2037	USU02687CJ55	15-Nov-2011	Issued	6.82	--	256,161,000	13,000	Regulation S	Senior Unsecured
AIG 6.820 15-Nov-2037	15-Nov-2037	US026874CW57	25-Sep-2012	Issued	6.82	026874CW5	256,133,000	143,369,000	Exchange Offer	Senior Unsecured
AIGVH / AIG 8.875 26-Jan-2040	26-Jan-2040	US91915WAB81	26-Jan-2010	Issued	8.875	91915WAB8	250,000,000	199,115,000	Underwritten	Senior Unsecured
AIGCFN / AIG 4.350 05-Apr-2042 '41	05-Apr-2042	US21871XAJ81	05-Apr-2022	Issued	4.35	21871XAJ8	500,000,000	500,000,000	Private placement-144A ...	Senior Unsecured
AIGCFN / AIG 4.350 05-Apr-2042 '41	05-Apr-2042	USU20256AE06	05-Apr-2022	Issued	4.35	--	500,000,000	500,000,000	Regulation S	Senior Unsecured
AIGFPM / AIG 0.900 17-Jun-2042 FRN MTN					0.9	--	5,291,000	5,291,000	Underwritten-Agent	Senior Unsecured
AIG 4.500 16-Jul-2044 '44	16-Jul-2044	US026874DA29	16-Jul-2014	Re-Opening	4.5	026874DA2	2,250,000,000	746,612,000	Underwritten	Senior Unsecured
AIG 4.800 10-Jul-2045 '45	10-Jul-2045	US026874DF16	10-Jul-2015	Issued	4.8	026874DF1	750,000,000	750,000,000	Underwritten	Senior Unsecured
AIGSAA / AIG 7.570 01-Dec-2045	01-Dec-2045	US00138GAB59	11-Jul-2013	Issued	7.57	00138GAB5	500,000,000	31,465,000	Private placement-144A ...	Junior Subordinated Uns...
AIGSAA / AIG 8.120 15-Mar-2046	15-Mar-2046	US026351AX43	15-Mar-1997	Issued	8.12	026351AX4	500,000,000	500,000,000	Private placement-144A	Senior Unsecured
AIGSAA / AIG 8.125 15-Mar-2046	15-Mar-2046	US00138GAC33	11-Jul-2013	Issued	8.125	00138GAC3	500,000,000	141,985,000	Private placement-144A	Junior Subordinated Uns...
AIG 4.576 27-Nov-2046 '36 MTN	27-Nov-2046	US02687QBF46	21-Nov-2006	Issued	4.5763	02687QBF4	20,000,000	18,246,000	Underwritten-Principal	Senior Unsecured
AIG 4.515 05-Dec-2046 '36 MTN	05-Dec-2046	US02687QBG29	05-Dec-2006	Re-Opening	4.515	02687QBG2	51,551,000	300,000	Underwritten-Principal	Senior Unsecured
AIG 4.511 01-Jun-2047 '37 MTN	01-Jun-2047	US02687QBX51	01-Jun-2007	Issued	4.5106	02687QBX5	39,055,000	39,055,000	Underwritten-Agent	Senior Unsecured
AIG 22-Nov-2047 '22	22-Nov-2047	XS1684198648	22-Nov-2017	Issued	0	--	400,000,000	400,000,000	Regulation S	Senior Unsecured
AIG 4.486 27-Nov-2047 '37 MTN	27-Nov-2047	US02687QDC96	27-Nov-2007	Re-Opening	4.4863	02687QDC9	67,471,000	2,500,000	Underwritten-Principal	Senior Unsecured

Figure A2: Some outstanding bonds of AIG

An example of coupled securities

Figure A2, taken from Refinitiv workspace, illustrates the common cases involving coupled securities for *American International Group*, henceforth AIG, as an example. Firstly, AIG has standalone Reg S (Maturity = 22 Nov 2047) and Rule 144A (maturity = 15 Mar 2046) bonds whose par values are independently recorded. It has also several sets of combined Reg S and Rule 144A bonds (for instance, the two with maturity = 05 Apr 2042) whose par values are calculated collectively together, as you can read from the tip shown beside the pointer in the photo. Finally, it has also tranches of three bonds, a reg S, a 144A, and a public bond (the three bonds with maturity = 15 Nov 2037). In this case, the two reg S and 144A were issued in 2011 with a par amount of 256,161,000 USD. Then, in 2012, 256,133,000 USD of the total par was exchanged for public bonds, and 13,000 USD remained as it was. The announcement for this change was as follows:

25-Sep-2012	Asset	USD 256,133,000 exchanged to CUSIP 026874CW5 on 25-Sep-2012. AIG is offering to exchange these notes due November 15, 2037 for its outstanding Series A-1 and Series A-6 Junior Subordinated Debentures.

References

Aghion, Philippe, and Patrick Bolton, 1992, An incomplete contracts approach to financial contracting, *Review of Economic Studies* 59, 473–494.

Aminadav, Gur, and Elias Papaioannou, 2020, Corporate Control around the World, *Journal of Finance* 75, 1191–1246.

Araujo, Aloisio P, Rafael VX Ferreira, and Bruno Funchal, 2012, The brazilian bankruptcy law experience, *Journal of Corporate Finance* 18, 994–1004.

Assunção, Juliano J, Efraim Benmelech, and Fernando SS Silva, 2014, Repossession and the democratization of credit, *Review of Financial Studies* 27, 2661–2689.

Bae, Kee-Hong, and Vidhan K. Goyal, 2009, Creditor Rights, Enforcement, and Bank Loans, *Journal of Finance* 64, 823–860.

Beck, Thorsten, Asli Demirg´uç-Kunt, and Vojislav Maksimovic, 2008, Financing patterns around the world: Are small firms different?, *Journal of Financial Economics* 89, 467–487.

Bolton, Patrick, and David S. Scharfstein, 1996, Optimal Debt Structure and the Number of Creditors, *Journal of Political Economy* 104, 1–25.

Brealey, Richard A., Stewart C. Myers, and Franklin Allen, 2020, *Principles of corporate finance*, thirteenth edition (McGraw-Hill Education, New York, NY).

Bretscher, Lorenzo, Lukas Schmid, Ishita Sen, and Varun Sharma, forthcoming, Institutional Corporate Bond Pricing, *Review of Financial Studies* .

Cerqueiro, Geraldo, Steven Ongena, and Kasper Roszbach, 2016, Collateralization, bank loan rates, and monitoring, *Journal of Finance* 71, 1295–1322.

Chava, Sudheer, and Michael R. Roberts, 2008, How Does Financing Impact Investment? The Role of Debt Covenants, *Journal of Finance* 63, 2085–2121.

Colla, Paolo, Filippo Ippolito, and Kai Li, 2013, Debt Specialization, *Journal of Finance* 68, 2117–2141.

Colla, Paolo, Filippo Ippolito, and Kai Li, 2020, Debt structure, *Annual Review of Financial Economics* 12, 193–215.

Demirgüç-Kunt, Asli, and Vojislav Maksimovic, 1999, Institutions, financial markets, and firm debt maturity, *Journal of Financial Economics* 54, 295–336.

Detragiache, Enrica, Paolo Garella, and Luigi Guiso, 2000, Multiple versus single banking relationships: Theory and evidence, *Journal of Finance* 55, 1133–1161.

Diamond, Douglas W., 2004, Presidential Address, Committing to Commit: Short-term Debt When Enforcement Is Costly, *Journal of Finance* 59, 1447–1479.

Djankov, Simeon, Oliver Hart, Caralee McLiesh, and Andrei Shleifer, 2008a, Debt enforcement around the world, *Journal of Political Economy* 116, 1105–1149.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer, 2008b, The law and economics of self-dealing, *Journal of Financial Economics* 88, 430–465.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer, 2007, Private credit in 129 countries, *Journal of Financial Economics* 84, 299–329.

Ellias, Jared E, and Elisabeth De Fontenay, 2025, The credit markets go dark, *The Yale Law Journal* 134, 779–857.

Esty, Benjamin C., and William L. Megginson, 2003, Creditor Rights, Enforcement, and Debt Ownership Structure: Evidence from the Global Syndicated Loan Market, *Journal of Financial and Quantitative Analysis* 38, 37–60.

Fan, Joseph PH, Sheridan Titman, and Garry Twite, 2012, An international comparison of capital structure and debt maturity choices, *Journal of Financial and quantitative Analysis* 47, 23–56.

Franks, Julian, Colin Mayer, and Stefano Rossi, 2009, Ownership: Evolution and Regulation, *Review of Financial Studies* 22, 4009–4056.

Gennaioli, Nicola, and Stefano Rossi, 2013, Contractual Resolutions of Financial Distress, *Review of Financial Studies* 26, 602–634.

Giannetti, Mariassunta, 2003, Do better institutions mitigate agency problems? evidence from corporate finance choices, *Journal of Financial and Quantitative Analysis* 38, 185–212.

Hackbarth, Dirk, Rainer Haselmann, and David Schoenherr, 2015, Financial distress, stock returns, and the 1978 bankruptcy reform act, *Review of Financial Studies* 28, 1810–1847.

Hart, Oliver, and John Moore, 1994, A theory of debt based on the inalienability of human capital, *Quarterly Journal of Economics* 109, 841–879.

Hart, Oliver, and John Moore, 1998, Default and renegotiation: A dynamic model of debt, *Quarterly Journal of Economics* 113, 1–41.

Hébert, Benjamin, and Jesse Schreger, 2017, The costs of sovereign default: Evidence from argentina, *American Economic Review* 107, 3119–3145.

John, Kose, Mahsa S. Kaviani, Lawrence Kryzanowski, and Hosein Maleki, 2017, Corporate debt maturity around the world: Role of creditor rights and contract enforcement, *Working Paper* .

John, Kose, Mahsa S Kaviani, Lawrence Kryzanowski, and Hosein Maleki, 2021, Do Country-Level Creditor Protections Affect Firm-Level Debt Structure Concentration?, *Review of Finance* 25, 1677–1725.

Kornejew, Martin, 2024, Insiders v. Outsiders: Market Creditor Rights, *Working Paper* .

La Porta, Rafael, Florencio Lopez-de Silanes, and Andrei Shleifer, 1999, Corporate Ownership around the World, *Journal of Finance* 54, 471–517.

La Porta, Rafael, Florencio Lopez-de Silanes, and Andrei Shleifer, 2008, The Economic Consequences of Legal Origins, *Journal of Economic Literature* 46, 285–332.

La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert W. Vishny, 1997, Legal Determinants of External Finance, *Journal of Finance* 52, 1131–1150.

La Porta, Rafael, Florencio Lopez-de Silanes, Andrei Shleifer, and Robert W. Vishny, 1998, Law and Finance, *Journal of Political Economy* 106, 1113–1155.

Li, Jian, and Haiyue Yu, 2022, Investor concentration, liquidity and bond price dynamics, *Working Paper* .

Li, Yingqi, Usha R. Mittoo, Xin Yu, and Zhou Zhang, 2021, The impact of global financial crisis on regulation S private debt market, *International Review of Economics & Finance* 73, 231–248.

Lilienfeld-Toal, Ulf von, Dilip Mookherjee, and Sujata Visaria, 2012, The distributive impact of reforms in credit enforcement: Evidence from indian debt recovery tribunals, *Econometrica* 80, 497–558.

Maggiori, Matteo, Brent Neiman, and Jesse Schreger, 2024, Corporate debt structure with home and international currency bias, *IMF Economic Review* 1–24.

Manconi, Alberto, Massimo Massa, and Lei Zhang, 2016, Bondholder Concentration and Credit Risk: Evidence from a Natural Experiment, *Review of Finance* 20, 127–159.

Mellor, Thomas, Marcus Marsh, and Suzanne Dabage De La Espriella, 2024, *An Introduction to Legal Risk and Structuring Cross-Border Lending Transactions*, 12th edition (Global Legal Group Ltd., London).

Ongena, Steven, and David C. Smith, 2000, What determines the number of bank relationships? cross-country evidence, *Journal of Financial Intermediation* 9, 26–56.

Ponticelli, Jacopo, and Leonardo S Alencar, 2016, Court enforcement, bank loans, and firm investment: evidence from a bankruptcy reform in brazil, *Quarterly Journal of Economics* 131, 1365–1413.

Qian, Jun, and Philip E. Strahan, 2007, How laws and institutions shape financial contracts: The case of bank loans, *Journal of Finance* 62, 2803–2834.

Rodano, Giacomo, Nicolas Serrano-Velarde, and Emanuele Tarantino, 2016, Bankruptcy law and bank financing, *Journal of Financial Economics* 120, 363–382.

Scott, Jonathan A, and Terence C Smith, 1986, The effect of the bankruptcy reform act of 1978 on small business loan pricing, *Journal of Financial Economics* 16, 119–140.

Thiel, Stuart E, 2000, Choice of law and the home-court advantage: evidence, *American Law and Economics Review* 2, 291–317.

Townsend, Robert M., 1979, Optimal contracts and competitive markets with costly state verification, *Journal of Economic Theory* 21, 265–293.

Vig, Vikrant, 2013, Access to collateral and corporate debt structure: Evidence from a natural experiment, *Journal of Finance* 68, 881–928.