

ESSAYS ON INDEX OWNERSHIP AND MARKET MICROSTRUCTURE

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Economics

By

Lu Yu, B.S.

Washington, DC
May 5, 2026

Copyright © 2026 by Lu Yu
All Rights Reserved

**ESSAYS ON INDEX OWNERSHIP AND MARKET
MICROSTRUCTURE**

Lu Yu, B.S.

Thesis Advisors: James Angel, Ph.D. and Toshihiko Mukoyama, Ph.D.

ABSTRACT

This is the abstract, a brief summary of the contents of the entire thesis. It is limited to 350 words.

INDEX WORDS: indexing, price informativeness, liquidity, volatility, market
microstructure

DEDICATION

The Dedication is optional.

ACKNOWLEDGMENTS

The author would like to thank...

PREFACE

A preface is not an introduction, and most theses do not need them.

CONTENTS

ACKNOWLEDGMENTS	v
PREFACE	vi
CHAPTER	
1 How Much Indexing is Good for the Market?	1
1.1 Introduction	1
1.2 Data	6
1.3 Baseline Analysis	10
1.4 Causality Analysis	16
1.5 Nonlinearity Analysis	26
1.6 Conclusion	45
2 Heterogeneous Effects of Index Ownership: Evidence from Firm Size and Business Cycle Conditions	48
2.1 Introduction	48
2.2 Size-Sorted Analysis	51
2.3 Business Cycle Analysis	57
2.4 Size and Business-Cycle Interactions	63
2.5 Conclusion	68
APPENDIX	
A First Appendix	71
A.1 Russell Rank Construction	71
A.2 IV Robutness Check - Full Results	73
A.3 Quadratic IV R1000/R2000 - Full Results	77
A.4 More Tables	78
B Second Appendix	84
B.1 More Tables	84
BIBLIOGRAPHY	91

LIST OF FIGURES

1.1	Index Ownership and Market Quality over Time	9
1.2	2SLS Regression Results of R1000	20
1.3	2SLS Regression Results of R2000	22
1.4	Quadratic Panel Regression Results	28
1.5	Quadratic Analysis - Market Score	30
1.6	Quadratic IV Estimates	32
1.7	Spline Regression Results	37
1.8	Spline Regression Results	38
1.9	Spline Regression Results	42
1.10	Spline Regression Results	43
2.1	Size-sorted Regression Results	53
2.2	Size-sorted Regression Results	54
2.3	Size-sorted Regression Results	56
2.4	Business-cycle Regression Results	59
2.5	Business-cycle Regression Results	60
2.6	Business-cycle Regression Results	61
2.7	Size and Business-Cycle Interactions Results	65
2.8	Size and Business-Cycle Interactions Results	66
2.9	Size and Business-Cycle Interactions Results	67

LIST OF TABLES

1.1	Panel Regressions with Firm and Time Fixed Effects: Detailed Output	13
1.2	First-Stage F Statistics for Global IV Cutoffs	18
1.3	IV Regression Results: Russell 1000	19
1.4	IV Regression Results: Russell 2000	21
1.5	First-Stage F Statistics for Spline IV at Global Russell 2000 Cutoff	41
A.1	IV Robustness Grid: Russell 1000	75
A.2	IV Robustness Grid: Russell 2000	76
A.3	Quadratic IV Comparison: Russell 1000	79
A.4	Quadratic IV Comparison: Russell 2000	80
A.5	Quadratic Nonlinearity Results	81
A.6	Spline Panel Regression Results	82
A.7	IV Regression Results for Spline IV at Global Russell 2000 Cutoff	83
B.1	Size-Sorted Regression Results by Market Capitalization Decile	85
B.3	Heterogeneity by Recession: Panel FE with Interaction	86
B.4	Business-Cycle by Size-Decile Regression Results: Amihud illiquidity	87
B.5	Business-Cycle by Size-Decile Regression Results: Volatility	88
B.6	Business-Cycle by Size-Decile Regression Results: Price informativeness	89
B.7	Best Business-Cycle by Size-Decile Results	90

CHAPTER 1

HOW MUCH INDEXING IS GOOD FOR THE MARKET?

1.1 INTRODUCTION

The rise of index investing has reshaped the composition of equity ownership and raised new questions about market quality. The growth of index investing is remarkable in the U.S. stock market. Introduction of the first index mutual funds was in the 1970s, followed by index ETFs in the 1990s. By the end of 2020, total net assets of these two index funds had grown to \$9.96 trillion, which is the total net asset value of the 2010 fund market. In the sample studied in this chapter, average index ownership increase from low single digits in the late 1990s to roughly 11~12% by 2023, indicating a broad-based shift rather than a tail-driven phenomenon.

Traditional finance theory has emphasized that active investors shape price informativeness, volatility, and liquidity through information production, monitoring, and trading. Much of the market-quality study therefore focuses on how informed trading and active monitoring shape price discovery and trading costs. The rise of passive ownership changes this benchmark by reallocating ownership toward index funds with limited incentives to gather firm-specific information. Understanding how index ownership affects market quality is thus a natural extension of the active-investor paradigm and central to modern asset markets.

While passive vehicles expand access to diversified portfolios and reduce trading costs, they may also alter incentives for information production and liquidity provision. As passive ownership grows, a larger share of trading can be driven by index flows rather than firm-specific signals, potentially affecting price discovery and volatility. These competing forces make it unclear whether greater index ownership improves or impairs liquidity, volatility, and price informativeness. Taken together, these facts motivate a systematic investigation of how index ownership affects liquidity, volatility, and price informativeness.

1.1.1 RELATED LITERATURE

A large literature emphasizes the role of active investors in producing and incorporating information into asset prices. In equilibrium, informed traders must earn rents to compensate for information acquisition (Grossman and Stiglitz, 1980), and price formation models describe how private information is impounded through trading (Kyle, 1985; Glosten and Milgrom, 1985). Market microstructure theory links asymmetric information to bid–ask spreads and liquidity, highlighting the importance of trading intensity and informed participation (Amihud and Mendelson, 1986; O’Hara, 1995). A related literature shows that arbitrage corrects mispricing but is limited by capital and risk constraints (Shleifer and Vishny, 1997). This active-investor benchmark provides the natural reference point for evaluating the rise of passive ownership.

A growing literature studies how passive investing affects market quality, with evidence pointing in both positive and negative directions. On the positive side, passive vehicles can deepen liquidity and reduce transaction costs by supplying predictable demand and supporting trading volume (Amihud and Mendelson, 1986; Admati and Pfleiderer, 1988). Empirical work documents that passive ownership can change trading activity and liquidity following index inclusion (Hjalmarsson, 2008) and that some

index additions improve market depth (Harris and Gurel, 1986; Shleifer, 1986). On the negative side, ETFs may transmit shocks across securities and increase return comovement and volatility (Ben-David, Franzoni, and Moussawi, 2018), while higher passive shares can reduce incentives for information production and weaken price efficiency (Grossman and Stiglitz, 1980; Petajisto, 2013; Kozak, Nagel, and Wang, 2020). These mixed findings motivate closer attention to the quasi-experimental evidence from index reconstitutions.

A large empirical literature examines index inclusion and demand-driven price effects. Early studies document price and volume changes around index additions, consistent with index fund demand shocks (Harris and Gurel, 1986; Shleifer, 1986). Subsequent research shows that arbitrage only partially offsets these shocks, implying persistent price impacts and downward-sloping demand curves for stocks (Wurgler and Zhuravskaya, 2002). Evidence from index revisions further shows that inclusion affects liquidity and trading behavior, reinforcing the role of passive flows in shaping market dynamics (Hjalmarsson, 2008). These findings motivate using index reconstitution as a quasi-experimental source of exogenous variation in passive ownership.

Taken together, the literature highlights both benefits and costs of passive ownership, but it provides limited guidance on the overall equilibrium effects on market quality. On the one hand, more index trading can lead to liquidity gain by reduced trading costs. On the other hand, a rising passive ownership may weaken incentives for information production and price discovery. These opposing channels imply that the relationship between indexing and market quality may be non-monotonic, and beyond some level the decline in information production could dominate the liquidity gains. As a result, a central open question is whether there exists an interior level of indexing that balances these forces.

The core research question is that how much indexing is good for the market. Market quality is measured using three complementary proxies: Amihud illiquidity, realized volatility, and a synchronicity-based price informativeness measure. A related question is whether the effect is monotonic or varies with ownership intensity. Over the same horizon of increasing index ownership, market-quality measures exhibit pronounced cyclical dynamics, which complicates inference. Amihud illiquidity spikes in recessions while the median remains near zero, suggesting that stress episodes are driven by a tail of illiquid observations. Volatility rises sharply in downturns and falls in expansions, and price informativeness dips during recessions before rebounding in recoveries. These patterns imply that the effects of indexing may be state-dependent and non-monotonic.

1.1.2 EMPIRICAL STRATEGY AND CONTRIBUTION

This paper contributes to the literature on passive investing and market quality by providing new empirical evidence on how index ownership shapes the functioning of equity markets. It examines the question using a long quarterly panel of 799,337 firm-quarter observations from 1993Q1 to 2023Q4 and a sequence of increasingly causal designs. The empirical strategy follows a standard progression from descriptive evidence to causal identification and functional-form testing.

First, baseline fixed-effects regressions document partial correlations between index ownership and each market-quality outcome. This provides a comprehensive analysis of how index ownership affects multiple dimensions of market quality—liquidity, volatility, and price informativeness—within a unified empirical framework. Prior work often isolates a single outcome or mechanism within a limited time horizon; in contrast, this study evaluates a broader set of market-quality measures and interprets them jointly. This design allows the analysis to assess whether improvements in

one dimension are accompanied by deterioration in another, which is central to the equilibrium tradeoff implied by passive ownership.

Second, an instrumental-variables design based on Russell 1000/2000 reconstitution provides exogenous variation in ownership to strengthen causal interpretation. This contributes to identification by exploiting exogenous variation in index ownership generated by Russell index reconstitution rules. The discontinuity at the Russell 1000/2000 cutoff induces sharp changes in index weights for otherwise similar firms, providing a quasi-experimental setting to estimate causal effects. This strategy strengthens the inference relative to purely cross-sectional correlations and directly links changes in index assignment to changes in market outcomes.

Third, quadratic specifications address the non-monotonic effects, testing whether marginal effects vary with ownership intensity, and identify potential interior optima. This explicitly models nonlinearities to test whether marginal effects vary across the ownership distribution and whether an interior optimum exists. To account for endogeneity, the nonlinear analysis is extended to an IV setting by instrumenting both ownership and its square to allow curvature in causal effects. The prevailing empirical literature largely focuses on average linear effects, even though theory predicts that passive investing can improve liquidity while weakening incentives for information production. By estimating specifications that allow the marginal effect of index ownership to vary with its level, the analysis evaluates whether these opposing mechanisms generate diminishing or adverse effects at high levels of indexing.

This chapter proceeds as follows. Section 1.2 presents the data construction and summary statistics that define the quarterly panel and core market-quality measures. Section 1.3 presents the baseline panel regression results and establishes the descriptive association between index ownership and market quality. Section 1.4 implements the IV strategy using Russell reconstitution to estimate causal effects and

interpret differences across cutoffs. Section 1.5 introduces the nonlinearity analysis using quadratic fixed-effects models and a composite market-quality score. Section 1.6 extends the nonlinearity framework to IV, evaluating whether curvature persists under causal. Section 1.7 concludes.

1.2 DATA

1.2.1 DATA DESCRIPTIONS

This section documents the construction and empirical properties of the quarterly panel used in subsequent regressions. The analysis combines holdings-based ownership measures with market-quality variables derived from daily trading data and then aggregates to the firm-quarter level. The resulting panel includes 799,337 firm-quarter observations spanning 1993Q1 through 2023Q4, with 26,395 distinct CUSIPs and 124 quarter-end dates. Each observation is keyed by firm identifier and quarter end and contains a primary transaction-cost measure alongside controls for size, trading activity, liquidity, and ownership.

The sample is an unbalanced panel. Firms enter and exit over time, and variable coverage varies across ownership measures and market-quality proxies. For descriptive statistics, the reported moments are computed using non-missing and non-infinite values within each column. This approach preserves the broad panel structure while ensuring that summary statistics are not distorted by computational artifacts.

Two features of the data merit emphasis at the outset. First, the long time coverage allows the analysis to span multiple business cycles, including recession and recovery periods, which is essential for later heterogeneity tests. Second, the distribution of key variables is highly skewed, which implies that summary statistics and regression

estimates should be interpreted with attention to dispersion and tails rather than only to central tendencies.

Index ownership and institutional ownership are constructed from mutual fund holdings reported in Thomson Reuters Mutual Fund Holdings (S12), accessed through WRDS. These holdings data are observed at quarter-end frequency and provide direct measures of fund-level portfolio positions. They cover a broad set of domestic mutual funds and ETFs with long time-series availability. This coverage supports analysis of ownership changes over multiple business cycles.

Market-quality outcomes and firm controls are derived from the CRSP daily stock file, including bid and ask prices, trading volume, market capitalization, and return. The CRSP records are observed daily and then aggregated to the quarter level for alignment with ownership data. Daily variables are therefore summarized within quarter windows to construct firm-quarter measures. This aggregation reduces high-frequency noise while maintaining comparability across firms and time.

Daily CRSP data are converted into quarterly measures using within-quarter aggregation rules that match each construct. Amihud illiquidity measure is computed from daily absolute returns scaled by daily dollar volume and then averaged over trading days in the quarter. Volatility is computed as the standard deviation of daily returns within the quarter, yielding a realized volatility measure for each firm-quarter. The price informativeness proxy is constructed from daily return-based inputs and then summarized at the quarter level, preserving the sign and scale implied by its transformation.

With these transformations in place, the merged dataset combines a holdings-based ownership panel with a trading-based market-quality panel into a unified quarterly file. The alignment choice treats ownership as a quarter-end state variable while market quality is measured over the same quarter. This timing convention reduces am-

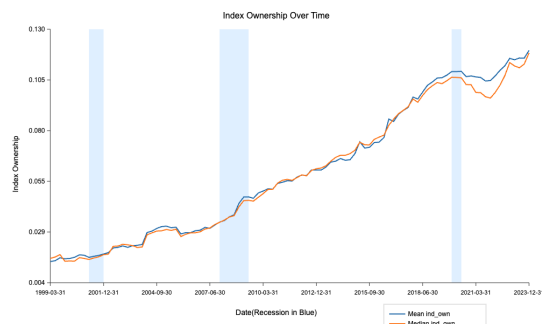
biguity about causal ordering in descriptive analyses. It also matches the fixed-effects framework used in later regressions.

1.2.2 KEY TRENDS

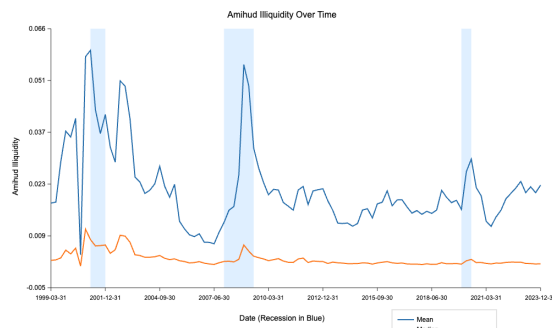
Liquidity and volatility measures are well-behaved in central tendency but display long upper tails. Amihud illiquidity measure has a mean of 0.136 and a median of 0.0128, a ratio that indicates heavy right skew typical of liquidity metrics. Realized volatility averages 0.182 with a median of 0.150, but the maximum exceeds 39, which likely reflects rare, high-stress quarters or firm-specific events. These tails motivate the use of fixed effects and robust inference in later regressions. Price informativeness displays a nearly symmetric spread around its mean, with a minimum of -6.91 and a maximum of 6.91. The interquartile range spans 0.385 to 2.571, indicating that most observations cluster around moderate informativeness values, while the tails capture extreme comovement or idiosyncratic dynamics.

Ownership variables are more concentrated. Index ownership averages 0.069 with an interquartile range from 0.026 to 0.095. Institutional ownership has a mean of 0.469 and a median of 0.476, indicating broad participation of institutions across the sample. The ownership distributions underscore that even modest changes in index holdings can be economically meaningful when they occur in large firms or in periods of heightened market stress.

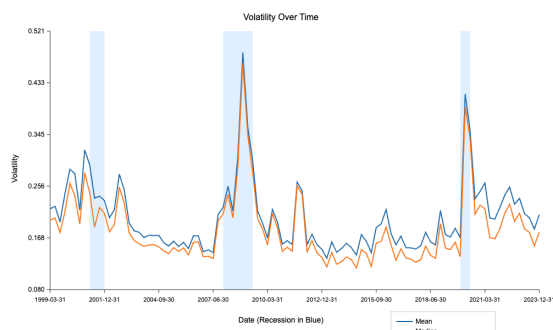
The constructed panel provides a long-horizon view of firm-level market quality, trading activity, and ownership structure. The data are rich in time coverage and cross-sectional breadth, yet they display notable skewness and ownership-related missingness. These features inform both the specification of baseline regressions and the interpretation of heterogeneity analyses in subsequent chapters.



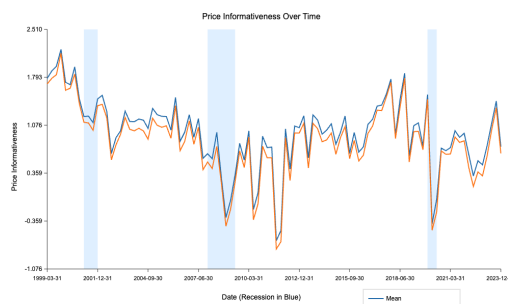
(a) Index Ownership



(b) Illiquidity



(c) Volatility



(d) Price Informativeness

Figure 1.1: Index Ownership and Market Quality over Time

Notes: Data Sources: WRDS-Thomson S12 and CRSP; Recessions are highlighted in blue.

Figure 1.1 show a steady increase in index ownership over the full sample, with mean and median ownership rising from low single digits in the late 1990s to roughly the low teens by 2023. Recession windows in 2001, 2007–2009, and 2020 are visible as brief slowdowns but do not reverse the long-run upward trend. The closeness of the mean and median lines indicates that the increase is broad-based rather than driven by a narrow tail of firms. This pattern suggests that index participation has deepened across the cross section even as macro conditions have fluctuated.

Liquidity, informativeness, and volatility move sharply with the cycle, but their behaviors differ in level shifts and distributional shape. Amihud illiquidity spikes in recessions while the median remains near zero, implying that stress episodes are driven by a tail of illiquid observations rather than a uniform shift in the cross section. The post-2009 decline in the mean suggests a gradual easing of liquidity frictions even as index ownership expands, which points to a coexistence of indexing growth and improved trading conditions outside stress windows. Price informativeness dips during recessions and then rebounds, and the close alignment of mean and median implies a broad distributional shift rather than a tail event. Volatility rises sharply in recessions and falls during expansions, with spikes that appear economy-wide rather than concentrated in a small subset of firms, indicating that macro shocks dominate firm-specific dispersion in those periods.

1.3 BASELINE ANALYSIS

1.3.1 PANEL REGRESSION MODEL

This section establishes the baseline panel specification that links index ownership to firm-level market-quality outcomes. The purpose is to provide a transparent descriptive benchmark that conditions on firm and time fixed effects within a long quarterly panel. The baseline regression lays the empirical foundation for the later identification and nonlinearity analysis. These estimates summarize partial correlations that are useful for comparison across outcomes and subsamples.

The baseline models are estimated separately for each outcome using a two-way fixed-effects design with two-way clustered standard errors. Firm fixed effects absorb time-invariant heterogeneity in trading frictions, while time fixed effects capture common macro and market-wide shocks. Standard errors are clustered by firm and

date to accommodate serial correlation within firms and cross-sectional dependence across firms in the same quarter. This specification is held constant across outcomes to ensure comparability of the index-ownership coefficient. It therefore provides a disciplined platform for subsequent extensions and robustness analysis.

The baseline panel regression model is constructed as:

$$Y_{i,q} = \alpha_i + \lambda_q + \beta \text{IndexOwn}_{i,q} + \gamma X'_{i,q} + \epsilon_{i,q} \quad (1.1)$$

where $Y_{i,q}$ is a proxy for market quality for stock i at quarter q , which includes the measurement for liquidity, volatility, price informativeness. $\text{IndexOwn}_{i,q}$ is the index ownership of stock i at quarter q the fraction of shares held by index funds. $X'_{i,q}$ is the control variable including $\ln(\text{MarketCap}_{i,q})$ and $\ln(\text{DollarVolume})$ controlling for firm size and trading intensity. α_i and λ_q are the firm and time fix effects.

The coefficient of interest is β , which captures the association between index ownership and each outcome. Because the specification includes firm and time fixed effects, β is identified from within-firm variation over time relative to common trends. This design reduces confounding from time-invariant firm heterogeneity and broad market shifts. Nevertheless, the estimates remain correlational because ownership may respond endogenously to market quality.

1.3.2 VARIABLE CONSTRUCTION

$\text{IndexOwn}_{i,q} = \frac{\text{IndShares}_{i,q}}{\text{Shrout}_{i,q}}$ is defined as the fraction of a firm's shares held by index funds at quarter end. This variable captures the intensity of passive ownership in the shareholder base. It is the primary explanatory variable in the baseline regressions. Higher values indicate greater exposure to indexing-related trading and information channels.

$$\text{Amihud}_{i,q} = \frac{1}{n} \sum_{t=1}^n \frac{|r_{i,t}|}{P_{i,t} * v_{i,t}}$$

Amihud illiquidity is constructed from daily data as the average of absolute returns divided by daily dollar volume within each quarter. The measure captures the price impact per unit of trading volume, so higher values indicate lower liquidity.

$$\text{Volatility}_{i,q} = \left(\sqrt{\frac{1}{N_q - 1} \sum_{d=1}^{N_q} (r_{i,d,q} - \bar{r}_{i,q})^2} \right) \times \sqrt{N_q}$$

Volatility is computed as the standard deviation of daily returns within each quarter, yielding a realized volatility measure. Both variables are first measured at daily frequency and then aggregated to the quarter to capture high-frequency liquidity variation while reducing idiosyncratic noise. This aggregation preserves temporal dynamics relevant for panel estimation and aligns the measures with quarter-end ownership data.

Price informativeness is proxied by a synchronicity-based measure defined as $\ln\left(\frac{1-R_{i,q}^2}{R_{i,q}^2}\right)$, where R^2 is obtained from a daily Fama–French three-factor regression as:

$$r_{it} - r_{ft} = \alpha_{it} + \beta_1 (r_{Mt} - r_{ft}) + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \epsilon_{it}$$

This approach is adopted because synchronicity captures the degree to which firm returns comove with common factors, a feature that reflects the incorporation of market-wide information into prices. Daily excess returns $r_{it} - r_{ft}$ are regressed on the market excess return, SMB, and HML within each quarter to compute firm-level R^2 , thereby retaining high-frequency information before aggregation. The quarterly transformation aggregates daily informativeness in a way that preserves economically meaningful variation across firms and time. The logarithmic transform maps

the bounded R^2 into an unbounded scale, improves comparability across firms, and yields coefficients that are more linear in interpretation.

Control variables include $\ln MarketCap_{i,q}$ and $\ln DollarVolume_{i,q}$ log of firm size and dollar trading volume, both computed from CRSP data and aggregated to the quarter. Firm size captures persistent differences in information environments, disclosure intensity, and investor attention that are correlated with both ownership structure and market quality. Dollar volume proxies for trading activity and liquidity provision, absorbing variation in market participation that could otherwise be attributed to index ownership. Including these controls reduces omitted-variable bias by isolating the partial association between index ownership and the outcome variables. The controls are standard in the market-quality literature (etc, Chordia, Roll, and Subrahmanyam, 2008; Chung and Hrazdil, 2010).

Table 1.1: Panel Regressions with Firm and Time Fixed Effects: Detailed Output

	Amihud illiquidity	Volatility	Price informativeness
Industry ownership	0.0613*** (0.0201)	-0.0560** (0.0233)	-3.3018*** (0.2384)
Firm size	0.0084*** (0.0016)	-0.0984*** (0.0031)	-0.3406*** (0.0269)
Dollar volume	-0.0222*** (0.0017)	0.0748*** (0.0024)	0.1453*** (0.0148)
Observations	241199	240515	239758
Entities	8877	8858	8855
Time periods	100	100	100
Firm FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Two-way clustered SE	Yes	Yes	Yes

Notes: This table reports panel fixed-effects regressions with two-way clustered standard errors by firm and date. Standard errors are reported underneath coefficients in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The detailed output preserves the model-level counts and specification flags used in the baseline panel regressions.

Table 1 reports the baseline coefficient estimates on index ownership for each outcome. The estimates are obtained from the two-way fixed-effects specification with two-way clustered standard errors, using the ownership-covered subsample. Each column is interpreted as a partial correlation between index ownership and the relevant market-quality proxy. The paragraphs below summarize the magnitude, sign, and statistical precision for each variable.

For Amihud illiquidity, the coefficient on index ownership is positive and statistically significant, with $\beta = 0.0613$ and a 95% confidence interval of $[0.0219, 0.1008]$. This estimate implies that higher index ownership is associated with higher illiquidity after conditioning on fixed effects and controls. The magnitude is economically meaningful because illiquidity exhibits a heavy right tail, so modest shifts in the mean can correspond to sizable changes in price impact for the most illiquid observations. In addition, the Amihud measure scales returns by dollar volume, which makes it particularly sensitive to changes in trading conditions during periods of thin liquidity. The result is significant at the 1% level, indicating a robust association in the baseline design.

For volatility, the coefficient on index ownership is negative and statistically significant, with $\beta = -0.0560$ and a 95% confidence interval of $[-0.1018, -0.0103]$. This estimate suggests that higher index ownership is associated with lower realized volatility in the baseline specification. The sign contrasts with concerns that passive ownership mechanically amplifies volatility, which motivates further regime and robustness analysis. The estimate is significant at the 5% level and remains stable within the ownership-covered sample.

For price informativeness, the coefficient is large, negative, and highly significant, with $\beta = -3.3018$ and a 95% confidence interval of $[-3.7691, -2.8345]$. This implies that higher index ownership is associated with lower values of the informativeness

proxy in the baseline panel. The magnitude is sizable relative to the cross-sectional dispersion of the proxy, indicating a strong negative association. This result is statistically significant at the 1% level and warrants additional scrutiny of measurement and identification in later chapters.

Control variables behave in expected directions across outcomes. Firm size is positively associated with illiquidity but negatively associated with volatility and the informativeness proxy. Dollar trading volume is negatively associated with illiquidity and positively associated with both volatility and informativeness. These patterns are consistent with larger and more actively traded firms exhibiting lower transaction costs but also different volatility and information dynamics.

The consistency of control coefficients across outcomes suggests that the model captures stable cross-sectional patterns. However, these controls do not resolve the endogeneity of index ownership. The baseline results therefore serve as descriptive benchmarks rather than causal estimates. Later identification strategies will be required to isolate causal effects of index ownership on market quality.

The baseline panel regressions establish three principal results: index ownership is positively associated with illiquidity, negatively associated with volatility, and negatively associated with the price informativeness proxy. These patterns are estimated within a two-way fixed-effects framework and provide a coherent descriptive benchmark across outcomes. At the same time, the mixed signs suggest that indexing may operate through multiple channels rather than a single monotonic mechanism. This motivates a deeper examination of identification and functional form in the chapters that follow.

A central limitation of the baseline results is endogeneity, because index ownership is likely jointly determined with market quality and may respond to unobserved shocks to liquidity, information, or volatility. Such feedback can bias fixed-effects estimates

even after controlling for firm and time effects. The next section therefore implements instrumental-variable regressions to isolate variation in index ownership that is plausibly exogenous. Beyond endogeneity, the mixed signs also raise the possibility that effects are nonlinear, with different responses at low versus high ownership intensity. Subsequent sections accordingly examine nonlinear and nonparametric specifications to test whether the baseline correlations vary across the ownership distribution.

1.4 CAUSALITY ANALYSIS

1.4.1 IDENTIFICATION

This chapter estimates the causal effect of index ownership on market-quality outcomes using instrumental variables. The objective is to address endogeneity that arises when ownership responds to unobserved shocks to liquidity, volatility, or information, or when these outcomes influence index inclusion probabilities. A two-stage least squares design is therefore used to isolate exogenous variation in index ownership. The IV estimates complement the baseline fixed-effects results by strengthening causal interpretation.

The Russell 1000/2000 reconstitution is a natural instrument because it generates a discrete, externally imposed change in passive ownership for firms around the cutoff. The ranking rule is mechanical and based on a pre-announced market-capitalization snapshot, so firms just above and below the threshold are similar *ex ante*. In this setting, assignment to the Russell 2000 produces a larger ownership shock than assignment to the Russell 1000 because small-cap index funds replicate more completely and cannot concentrate in a few mega-cap names. This design follows the identification logic used in studies of passive ownership and index reconstitution (Ben-David et

al., 2018; Chang, Hong, and Liskovich, 2015; Appel, Gormley, and Keim, 2016; Crane et al., 2016).

The instrument variables is valid as it satisfies the relevance and exclusion conditions. Relevance holds because assignment to the Russell 2000 causes a discrete jump in passive ownership, which is visible in the first stage.

$$\text{Cov}(Z_{i,q}, \text{IndexOwn}_{i,q}) \neq 0$$

The exclusion restriction requires that Russell reconstitution affects market quality only through its impact on index ownership.

$$Z_{i,q} \perp \varepsilon_{i,q}$$

Firms around the cutoff in the Russell reconstitution are similar in underlying fundamentals, and inclusion in Russell 1000/2000 itself does not directly change firm characteristics or trading environment. Therefore conditional on controls for firm size and fixed effects, the variation in index ownership induced by the Russell reconstitution provides plausible exogenous variation for identifying the causality.

1.4.2 2SLS REGRESSION SETUP

First Stage:

$$\text{IndexOwn}_{i,q} = \pi_0 + \pi_1 Z_{i,q} + \gamma X'_{i,q} + \alpha_i + \lambda_q + u_{i,q}$$

where

$$Z_{i,q} = 1 \{ \text{Rank}_{i,q} \geq c \}, \tag{1.2}$$

$c =$ choice of cutoffs, ($R1000$ or $R2000$)

Second Stage:

$$Y_{i,t} = \beta_0 + \beta_1 \widehat{\text{IndexOwn}}_{i,q} + \gamma X'_{i,q} + \alpha_i + \lambda_q + \epsilon_{i,q} \tag{1.3}$$

Tables 2 reports first-stage diagnostics for the global IV specifications at the Russell 1000 and Russell 2000 cutoffs. Using the full sample with a sharp indicator yields strong first-stage F statistics: approximately 28 at the 1000 cutoff and above 128 at the 2000 cutoff. These values indicate that the instrument is strong in the global specification, producing precise fitted ownership variation. The stronger first stage at the 2000 cutoff reflects a larger discontinuity in passive ownership for firms entering the small-cap index.

Table 1.2: First-Stage F Statistics for Global IV Cutoffs

	Russell 1000	Russell 2000
Amihud illiquidity	27.7911 (p=1.35e-07)	131.1626 (p=<1e-16)
Volatility	27.3002 (p=1.74e-07)	128.9754 (p=<1e-16)
Price informativeness	28.4864 (p=9.44e-08)	131.1246 (p=<1e-16)
Observations	235544	235544
Firms	8627	8627
Dates	99	99
Clusters	8627	8627
Controls	Yes	Yes
Fixed effects	cusip FE + date FE	cusip FE + date FE
SE type	firm clustered	firm clustered

Notes: This table reports first-stage relevance diagnostics for the global cutoff-IV design. Entries are first-stage F statistics for the excluded cutoff indicator in the first-stage regression of instrumented ownership. P-values are shown in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The sample is the global specification with no bandwidth restriction, estimated separately at the Russell 1000 and Russell 2000 cutoffs.

Local RD-IV estimates provide a credibility check but exhibit weak first stages in narrow bandwidths. Within a ± 300 -rank window around the 2000 cutoff and a ± 150 window around the 1000 cutoff, first-stage F statistics fall well below conventional thresholds. This weakness is expected because the ownership discontinuity is smaller in very narrow windows and noise dominates local variation. Accordingly, detailed

local RD-IV results are reported as robustness check in the Appendix section, where the trade-off between identification purity and statistical power is examined explicitly.

The contrast between local and global IV results reflects a standard trade-off in reconstitution designs. Local RD-IV estimates are more credible with respect to the exclusion restriction but suffer from weak first stages due to limited ownership discontinuities near the threshold. Global specifications yield strong instruments and precise estimates but rely on the stronger assumption that potential outcomes vary smoothly across the full rank distribution. Following common practice, the global estimates are treated as primary results, while local estimates serve as robustness checks.

Table 1.3: IV Regression Results: Russell 1000

		Russell 1000
Amihud illiquidity		0.2319 (0.1450)
Volatility		-1.4850*** (0.4090)
Price informativeness		-5.8808 (4.0159)
Observations		235544
Firms		8627
Dates		99
Clusters		8627
Endogenous variable		<i>ind_own</i>
Instrument		<i>z_rank_le_cutoff</i>
Controls		<i>mktcap; c_dollar_vol; r_tno; rank_minus_cutoff</i>
Fixed effects		cusip FE + date FE
SE type		firm clustered

Notes: This table reports second-stage IV estimates for the global cutoff-IV design. Standard errors are reported underneath coefficients in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The sample is the global specification with no bandwidth restriction, estimated at the Russell 1000 cutoff.

Table 3 summarizes the global IV estimates at the Russell 1000 cutoff. Index ownership is positively associated with Amihud illiquidity ($\beta = 0.2319$, $p \approx 0.11$),

negatively associated with volatility ($\beta = -1.4850$, $p < 0.001$), and negatively associated with price informativeness ($\beta = -5.8808$, $p \approx 0.14$).

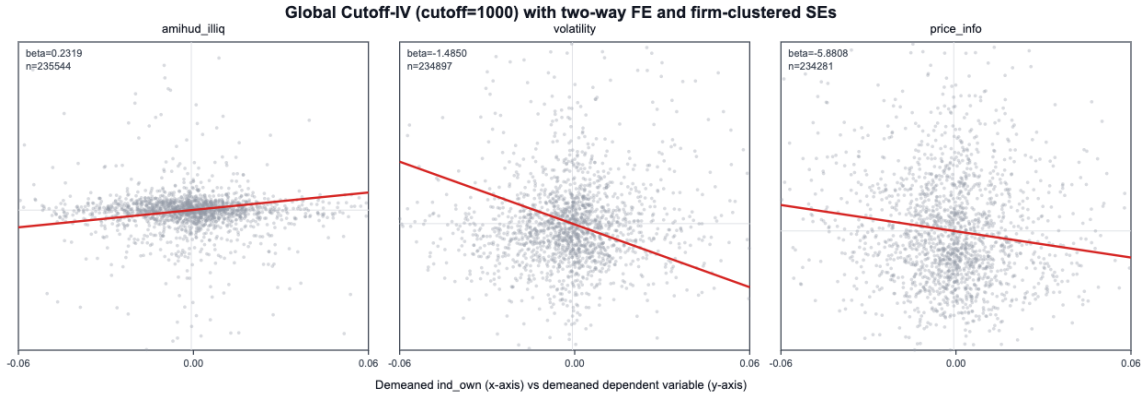


Figure 1.2: 2SLS Regression Results of R1000

*Notes:*Note: This figure plots fitted IV relationships between index ownership and each market-quality outcome using the global Russell cutoff 1000 specification.

Figure 1.2 presents the IV results at the Russell 1000 cutoff for all three market quality measures. For Amihud illiquidity, the upward line would imply a deteriorated liquidity in the more passive market. However, the estimate is not statistically significant. Therefore, the Russell 1000 IV results do not support a reliable liquidity conclusion. The volatility result is materially different. The fitted line in the volatility panel slopes downward in a clear and economically meaningful way. This indicates that higher index-related ownership reduces idiosyncratic volatility for firms near the large-cap boundary. For the price informativeness proxy, the estimated coefficient is also negative. The sign is directionally consistent with the baseline concern that passive ownership may weaken firm-specific price discovery, yet the estimate is too statistically imprecise to support a meaningful causal conclusion. Taken together, the cutoff-1000 plot points primarily to lower volatility, while the illiquidity and informativeness estimates are directionally consistent but less precise. The main implication of the Russell 1000 specification is therefore a volatility-reduction channel.

Table 1.4: IV Regression Results: Russell 2000

	Russell 2000
Amihud illiquidity	-0.2849*** (0.0776)
Volatility	-0.1767 (0.1263)
Price informativeness	-12.0868*** (1.7568)
Observations	235544
Firms	8627
Dates	99
Clusters	8627
Instrument	<i>z_rank_le_cutoff</i>
Controls	<i>mktcap; c_dollar_vol; r_tno; rank_minus_cutoff</i>
Fixed effects	cusip FE + date FE
SE type	firm clustered

Notes: This table reports second-stage IV estimates for the global cutoff-IV design. Standard errors are reported underneath coefficients in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The sample is the global specification with no bandwidth restriction, estimated at the Russell 2000 cutoff.

The global cutoff-2000 specification is substantially stronger from an identification standpoint. Its first-stage F-statistics are around 129 to 131, indicating a much more powerful relationship between index assignment and index ownership. This stronger first stage gives the second-stage coefficients greater credibility and makes the visual patterns in the fitted-line plots more informative.

Table 4 reports the global IV estimates at the Russell 2000 cutoff. Index ownership is associated with lower illiquidity ($\beta = -0.2849$, $p < 0.001$), indicating a sizable liquidity improvement for firms entering the small-cap index. The price informativeness coefficient is large and negative ($\beta = -12.0868$, $p < 10^{-11}$), suggesting that increased passive ownership is linked to lower values of the synchronicity-based proxy.

The volatility estimate is negative but statistically weak ($\beta = -0.1767$, $p \approx 0.16$), implying that volatility effects are less precisely identified at this cutoff. Taken together, the 2000-cutoff results point to a strong liquidity channel and a pronounced informativeness response, while volatility effects remain ambiguous.

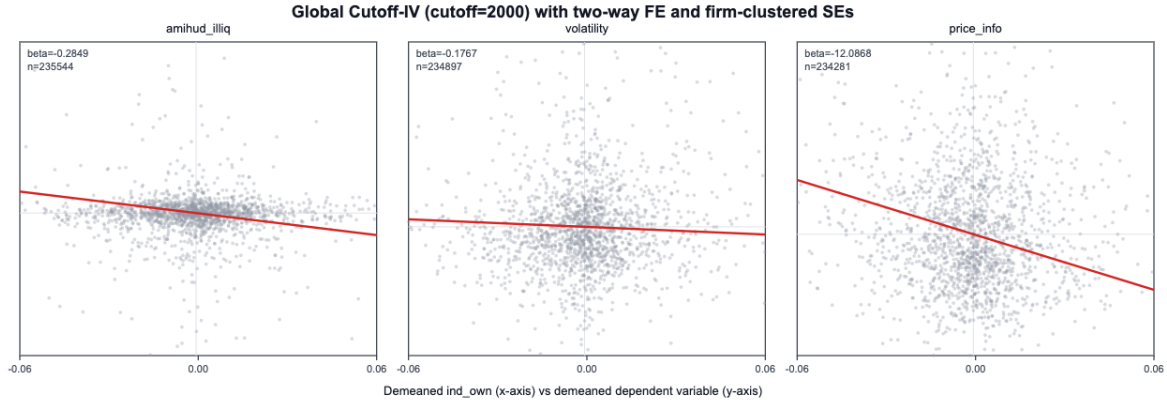


Figure 1.3: 2SLS Regression Results of R2000

*Notes:*Note: This figure plots fitted IV relationships between index ownership and each market-quality outcome using the preferred global Russell cutoff 2000 specification. .

In Figure 1.3, the fitted line in the illiquidity panel slopes downward in a manner that matches the regression result, showing the liquidity is improved in a market with higher indexing investment. Accordingly, the cutoff-2000 IV evidence suggests that passive ownership can improve trading conditions for smaller firms. The fitted line in the price-informativeness panel is clearly downward sloping and is visually the most persuasive panel among the cutoff-2000 plot. This result supports the view that passive ownership weakens firm-specific information production or incorporation into prices among smaller firms near the Russell 2000 boundary. In this specification, the information environment appears to be the main cost associated with index-driven ownership changes.

The fitted line in the volatility panel also slopes downward, but less decisively than the lines for illiquidity and informativeness. It is not statistically significant and

the evidence is not precise enough to support a firm conclusion. Therefore, the overall interpretation of the cutoff-2000 plot is that higher instrumented ownership improves liquidity while simultaneously reducing the informativeness proxy, with the volatility channel remaining comparatively weak.

1.4.3 WHY CUTOFFS PRODUCE DIFFERENT EFFECTS

There is an interesting observation that the two global cutoff specifications imply materially different causal effects. The estimated effects differ in both economic channel and statistical precision. At the Russell 1000 cutoff, the only robust result is the reduced-volatility channel, while the estimated effects on liquidity and informativeness are statistically weak. At the Russell 2000 cutoff, the volatility effect fades, but the liquidity and informativeness effects become prominent and statistically significant. This pattern indicates that the effect of passive ownership is not stable across the size distribution. Instead, the dominant adjustment margin depends on the type of firms exposed to the ownership shock.

Although the liquidity effect in the Russell 1000 sample is not statistically significant, this sign reversal is economically suggestive. Around the Russell 1000 boundary, the illiquidity coefficient is positive but imprecise, whereas around the Russell 2000 boundary it is negative and strongly significant. The Russell 1000 estimate should not be overstated because it lacks statistical precision, yet the sign reversal is still economically suggestive. It implies that passive ownership may interact differently with pre-existing trading environments in large-cap and small-cap stocks. In large-cap firms, additional passive ownership appears unable to generate clear liquidity gains, while in smaller firms it is associated with a measurable improvement in trading conditions.

The divergent patterns around the Russell 1000 and Russell 2000 cutoffs are consistent with systematic differences in firm characteristics and market structure. Firms near the Russell 1000 threshold are relatively large, trade in deeper markets, and receive broader attention from investors and intermediaries. In that environment, baseline liquidity and information production are already comparatively strong, so an increase in passive ownership is less likely to generate sizable marginal improvements in transaction costs or price efficiency. Instead, the effect of indexing is more likely to appear through trading-related channels, such as correlated fund flows and changes in return dynamics. Firms around the Russell 2000 cutoff, by contrast, are smaller, less liquid, and subject to greater informational frictions, so the same increase in index ownership represents a more meaningful expansion in market participation.

This difference in market structure helps explain why the Russell 2000 specification produces stronger effects on liquidity and the price informativeness proxy. When the initial investor base is thinner, passive inflows can increase trading activity and reduce trading frictions, so the ownership shock has a larger effect on market quality. By contrast, when those frictions are already limited, as for firms near the Russell 1000 cutoff, the incremental effect of passive ownership is correspondingly smaller and less likely to appear in standard liquidity measures. The contrast between the two cutoffs therefore points to heterogeneity in the causal effect of indexing rather than inconsistency in the empirical evidence.

Two mechanisms are particularly relevant. First, passive inflows may serve as a net addition to trading activity among smaller-cap stocks, where the marginal value of new participation is high. Second, substitution between active and passive investors is likely to be more pronounced among firms near the Russell 1000 cutoff, where passive products are already widely used and active traders continue to play an important role in price discovery. If passive ownership expands mainly by displacing active investors

in that segment, the gains in liquidity and information efficiency may be attenuated. In addition, because larger firms are more deeply embedded in exchange-traded funds and benchmark-tracking portfolios, which generate correlated trading flows. These flows can increase return comovement and amplify non-fundamental volatility without necessarily improving liquidity or price efficiency.,

Differences in treatment intensity and identification further support this interpretation. The discontinuity in index ownership is typically larger around the Russell 2000 cutoff than around the Russell 1000 cutoff, which yields a stronger first stage and more precise second-stage estimates. As a result, the Russell 2000 specification is better positioned to detect how passive ownership affects market quality when the underlying firms are more sensitive to capital flows. The two cutoffs therefore identify effects in distinct local environments: the Russell 1000 captures relatively liquid, information-rich large-cap settings, whereas the Russell 2000 captures smaller-cap settings in which market frictions are more substantial and the ownership shock is more economically meaningful.

Taken together, these findings imply that the causal effect of index ownership is context-dependent rather than uniform across the cross section. The two thresholds should therefore be interpreted as complementary evidence on how indexing operates across different firm size regimes rather than as competing empirical conclusions. This interpretation motivates the next stage of the analysis, which turns to size-sorted panel regressions in order to examine whether similar patterns appear more broadly across the full size distribution.

The IV analysis identifies the causal effect of index ownership using Russell reconstitution as an exogenous source of variation. In the global specifications, the first stage is strong at both cutoffs, confirming substantial ownership discontinuities. At the R1000 cutoff, index ownership is associated with lower volatility, while effects on

illiquidity and price informativeness are weaker and less precise. At the R2000 cutoff, index ownership is associated with a large improvement in liquidity and a sizable decline in the price informativeness proxy, with the volatility estimate negative but statistically weak. Overall, the IV evidence points to meaningful liquidity improvements for smaller firms and nontrivial informativeness effects, alongside a more robust volatility response at the large-cap boundary.

These findings raise the question of whether the ownership–market-quality relationship is nonlinear across the distribution of index ownership. If marginal effects differ between low- and high-ownership regimes, linear IV estimates may mask important heterogeneity. The next section therefore examines nonlinear and nonparametric specifications to test whether the causal effects vary with ownership intensity and firm characteristics. This extension clarifies whether the cutoff-specific differences reflect broader nonlinear patterns in the data.

1.5 NONLINEARITY ANALYSIS

1.5.1 QUADRATIC ANALYSIS SET-UP

This section extends the baseline panel and IV analyses by examining whether the relationship between index ownership and market quality is nonlinear. The linear specifications documented systematic effects but also produced mixed signs across outcomes and cutoffs, raising the possibility that marginal effects vary with ownership intensity. A nonlinear framework can reconcile these patterns by allowing the slope to change over the ownership distribution.

Exploring nonlinearity is economically motivated because the mechanisms of indexing need not scale proportionally. Small increases in passive ownership may improve liquidity and information aggregation, while very high ownership could dampen

trading incentives or concentrate demand. Such threshold effects would be masked in linear models but are consequential for policy and interpretation. The analysis therefore estimates quadratic specifications and constructs a composite market-quality score to identify an interior optimum.

The baseline nonlinear specification augments the fixed-effects model with a quadratic term:

$$Y_{i,q} = \alpha + \beta_1 \text{IndexOwn}_{i,q} + \beta_2 \text{IndexOwn}_{i,q}^2 + \gamma X'_{i,q} + \alpha_i + \lambda_q + \epsilon_{i,q} \quad (1.4)$$

where the turning point is defined as

$$\text{IndexOwn}^* = -\frac{\beta_1}{2\beta_2}$$

Figure 1.4 plots the fixed-effects quadratic fits for each outcome over the observed ownership support, normalized relative to the median ownership level.

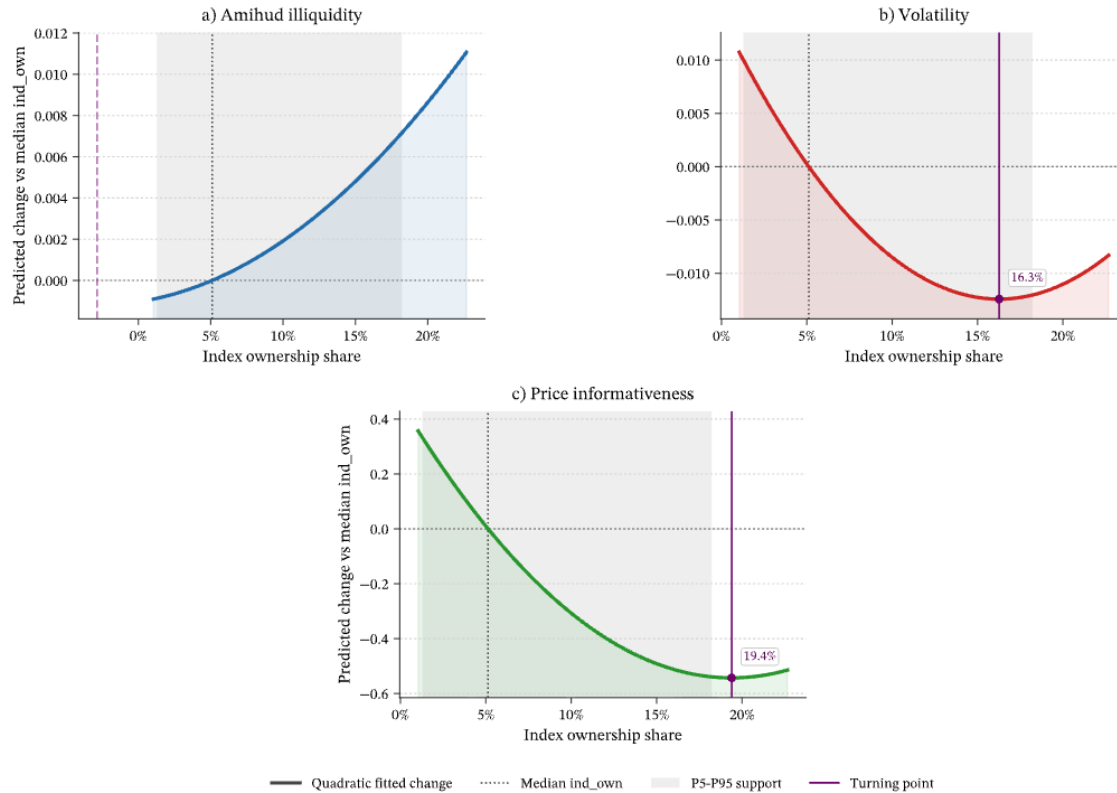


Figure 1.4: Quadratic Panel Regression Results

*Notes:*Note: This figure shows the non-IV quadratic relationship between index ownership and the three market-quality outcomes. Each panel traces the fitted nonlinear effect over the support of index ownership, illustrating whether the association between indexing and market quality is monotonic or exhibits an interior optimum.

Volatility exhibits a clear U-shape: marginal increases in index ownership reduce volatility at low levels, but the benefit diminishes and eventually reverses at higher ownership. Price informativeness displays a similar U-shape, implying that moderate indexing improves informativeness while very high ownership may dilute firm-specific signals. Amihud illiquidity shows weaker curvature, with the quadratic term significant but the turning point outside the observed support, suggesting a less stable liquidity response in the reduced-form setting.

These patterns imply that indexing is most beneficial at low-to-moderate levels, where passive participation can deepen trading and enhance information aggregation

without displacing active monitoring. As ownership rises, marginal benefits taper and may reverse, consistent with reduced incentives for information production or concentrated demand elasticity.

High levels of index ownership may be undesirable because they can reduce the incentives for costly information production and amplify demand inelasticity. When passive ownership is dominant, fewer investors trade on firm-specific signals, which can weaken price discovery and reduce informational trading depth. In addition, index-linked demand can concentrate order flow and increase co-movement, potentially elevating systemic sensitivity in stressed periods. These mechanisms provide an economic rationale for the observed reversal in marginal benefits at higher ownership levels.

1.5.2 COMPOSITE MARKET-QUALITY SCORE

The composite market-quality score aggregates the three quadratic outcome curves into a single metric with consistent economic signs. For each outcome, predicted changes over an `ind_own` grid are standardized to z-scores to make magnitudes comparable. The score is defined as

$$\text{score}(x) = -z_{\text{amihud}}(x) - z_{\text{volatility}}(x) + z_{\text{price_info}}(x) \quad (1.5)$$

reflecting that lower illiquidity and volatility are beneficial while higher informativeness is beneficial. This construction allows a unified assessment of market quality when outcomes move in different directions.

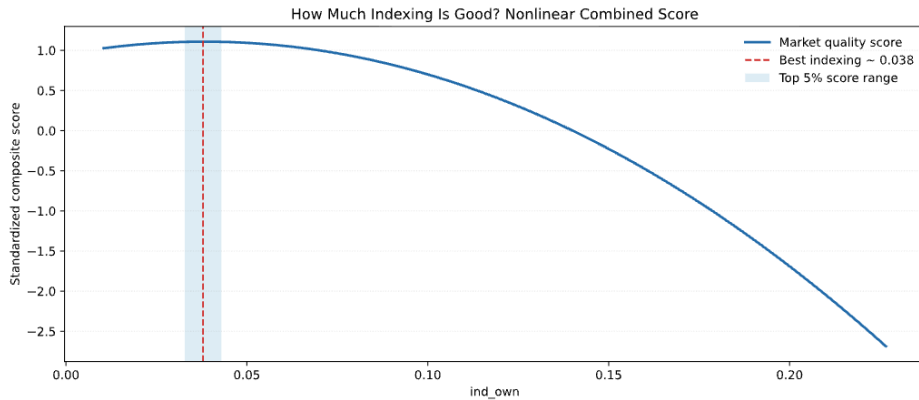


Figure 1.5: Quadratic Analysis - Market Score

*Notes:*Note: This figure reports the composite market-quality score implied by the non-IV quadratic specification. Higher values indicate a more favorable combination of higher liquidity, lower volatility, and higher price informativeness. The highlighted region marks the range of index ownership associated with the highest composite market quality.

Figure 1.5 plots the score over the ownership grid and highlights the ownership interval that falls in the top 5% of scores.. The score peaks at $\text{ind_own} \approx 0.0379$, with a top 5% range of approximately $[0.0328, 0.0430]$. This range implies that low-to-moderate indexing balances the liquidity and volatility benefits of passive participation with the information costs of high ownership. Low-to-moderate indexing level seems to be optimal for market quality under the composite criterion. The narrow high-quality range indicates that marginal effects change quickly around the optimum, so small deviations can reduce overall market quality. This result translates the nonlinear estimates into a concrete reference for policy-relevant statement about the level of indexing most consistent with market-quality improvement.

1.5.3 QUADRATIC IV ANALYSIS

The current design is reduced-form and may still reflect endogeneity in index ownership. To assess whether the nonlinear patterns persist under causal identification, the analysis extends the IV framework to a quadratic specification.

The Quadratic IV equation is defined as:

$$Y_{i,q} = \alpha + \beta_1 \widehat{\text{IndexOwn}}_{i,q} + \beta_2 \widehat{\text{IndexOwn}}_{i,q}^2 + \gamma X'_{i,q} + \alpha_i + \lambda_q + \epsilon_{i,q} \quad (1.6)$$

where $\widehat{\text{IndexOwn}}_{i,q}$ the predicted index ownership from the first-stage IV regression (instrumented by Russell 1000/2000 reconstitution cutoff).

The first stage is defined as:

$$\text{IndexOwn } n_{i,q} = \pi_0 + \pi_1 IV_{i,q} + \pi_2 IV_{i,q}^2 + \phi X'_{i,q} + \alpha_i + \lambda_q + u_{i,q} \quad (1.7)$$

Controls include firm size and dollar volume, along with the running variable and its square, while firm and time fixed effects are retained. This design parallels the linear IV strategy but permits curvature in the causal response.

The preferred IV specification uses the global cutoff at 2000, which provides strong first-stage diagnostics for both ind_own and ind_own^2 . Figure 1.6 presents the quadratic IV fitted lines under the preferred global Russell 2000 design.

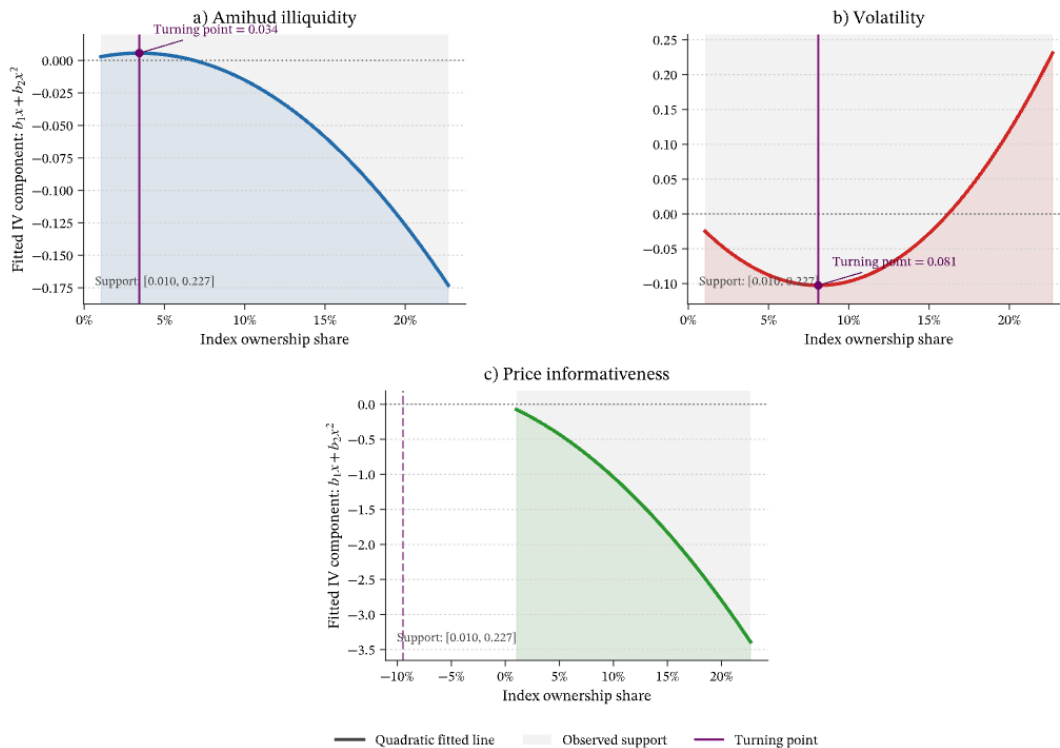


Figure 1.6: Quadratic IV Estimates

*Notes:*Note: This figure plots the fitted quadratic IV component for each market-quality outcome under the preferred global Russell cutoff 2000 specification. .

For volatility, the IV estimates display a clear U-shape, with $\beta_1 = -2.5336$ and $\beta_2 = 15.6492$ and a turning point around $\text{ind_own} \approx 0.0809$ within the support. This pattern implies that low-to-moderate indexing reduces volatility, while higher ownership levels are associated with diminishing and eventually reversing marginal benefits.

The liquidity and informativeness panels are less decisive. For Amihud illiquidity, the quadratic term is negative and statistically significant ($\beta_2 = -4.8236$), but β_1 is not significant, indicating weaker and less stable curvature. For price informativeness, the nonlinear terms are not statistically significant, suggesting limited IV evidence of curvature in that channel.

The cutoff-1000 quadratic IV design is not retained as a primary specification because the first-stage diagnostics become unstable and not economically meaningful; the corresponding results are reported as a robustness check in the Appendix section. This instability likely reflects weaker ownership discontinuities and a noisier instrument when nonlinear terms are introduced, which reduces power and inflates standard errors. In quadratic IV settings, instruments must be strong enough to identify both the level and the square of ownership, a requirement that the cutoff-1000 design does not satisfy. As a result, the 1000-cutoff estimates are treated as a failed robustness check rather than credible nonlinear evidence.

More broadly, quadratic IV estimation is demanding because it amplifies weak-instrument concerns and sensitivity to specification. Even with a strong first stage for `ind_own`, identification of `ind_own`² can be fragile if the instrument generates limited curvature in fitted values. This limits the precision of higher-order terms and explains why only the 2000-cutoff specification yields interpretable volatility curvature. These constraints motivate cautious interpretation and the use of multiple specifications to triangulate the nonlinear effects.

Despite these limitations, the quadratic IV results are economically meaningful in the channel where identification is strongest. The volatility U-shape implies that the marginal effect of indexing is beneficial at low-to-moderate ownership levels but becomes less favorable as ownership intensifies, which aligns with theories of reduced information production and demand inelasticity at high passive shares. The lack of robust curvature for price informativeness suggests either weaker causal effects or insufficient power to detect nonlinearities in that channel. Accordingly, the quadratic IV results should be interpreted as evidence of a volatility-specific nonlinear mechanism rather than a universal nonlinear effect across all market-quality dimensions.

This part shows that nonlinearity is a first-order feature of the relationship between index ownership and market quality. In reduced-form quadratic FE models, both volatility and price informativeness display U-shaped patterns, while illiquidity shows weaker curvature with an out-of-support turning point. A composite market-quality score aggregates these curves and indicates an interior optimum at $\text{ind_own} \approx 0.0379$, with a narrow high-quality range around 3.3%–4.3%. These results imply that low-to-moderate indexing tends to improve market quality, while very high ownership can erode marginal benefits.

The IV-based nonlinear analysis refines this conclusion by isolating causal curvature where identification is strongest. Using the cutoff-2000 design, volatility exhibits a robust U-shape with a turning point around $\text{ind_own} \approx 0.0809$, whereas illiquidity curvature is weaker and price informativeness shows no robust nonlinear response. The cutoff-1000 quadratic IV is not retained as a primary specification because its first-stage diagnostics are unstable; those results are reported as robustness checks in the appendix. Overall, the evidence indicates that nonlinear effects are most credible in the volatility channel, while other channels are less stable in IV space.

Taken together, the quadratic reduced-form and IV results show that indexing ownership is associated with market quality in a clearly nonlinear manner, but the evidence also suggests that a single global quadratic restriction is too coarse to summarize the full pattern. The estimated turning points are informative, yet they impose one smooth curvature over the entire support and therefore cannot distinguish whether the relationship changes gradually across the ownership range or shifts more sharply at specific intervals. This limitation is especially relevant because the volatility result is strongest at low-to-moderate ownership levels, while the liquidity and informativeness channels appear weaker and less stable. A spline specification is therefore the natural next step: it preserves a flexible nonlinear structure, allows the slope

to vary across segments, and provides a more credible basis for testing whether the apparent interior optimum is stable or only an artifact of the quadratic functional form.

1.5.4 SPLINE ANALYSIS

The spline analysis begins from the same question as the quadratic analysis: whether indexing ownership improves market quality, and if so, over what range of ownership the improvement is strongest. The earlier quadratic results suggest interior curvature, but a single global polynomial may be too restrictive to summarize the shape of data. A flexible spline is a more appropriate tool as it allows the slope to vary across the ownership distribution while preserving a smooth fit over the observed support. In other words, the spline specification tests whether the apparent optimum survives once the data are allowed to determine the local slope more freely.

In practice, the effects of index ownership on market functioning are likely to differ across regimes - between low, medium and high levels of ownership. The quadratic form assumes that marginal effects change smoothly and symmetrically around a unique point, which may mask the locally varying dynamics. By allowing curvature to vary across the support, the spline framework provides a piecewise-polynomial approximation of the relationship. It is better suited to trace that tradeoff and to show whether the estimated interior region is stable across reduced-form and IV specifications.

The spline specification re-estimates the relationship between index ownership and market quality using a natural cubic spline rather than a quadratic polynomial. The spline allows the slope to vary across the ownership distribution, which is useful when the benchmark quadratic suggests an interior optimum but may impose too much global smoothness. To keep the fit economically relevant, ownership is clipped

to the pooled $q_{01} - q_{99}$ support and rescaled to $[0,1]$. Three interior knots are placed at the 25th, 50th, and 75th percentiles, corresponding to raw values 0.0265, 0.0510, and 0.0952. Standard errors are two-way clustered by firm and date in the reduced-form panel model. The resulting estimates should be read as descriptive evidence on local curvature rather than as a causal design.

Let $IO_{i,q}$ to be index ownership and k be a knoe,

$$Y_{iq} = \beta_1 IO_{iq} + \sum_{k=1}^K \beta_{k+1} (IO_{iq} - \kappa_k)^+ + \gamma X'_{iq} + \alpha_i + \lambda_q + \varepsilon_{iq} \quad (1.8)$$

where

$$(IO_{it} - \kappa_k)^+ = \max(0, IO_{it} - \kappa_k), \quad k = 1, 2, 3, \dots$$

Stabilization is applied because the raw ownership distribution is highly skewed and contains a sparse upper tail. Clipping the series to the pooled $q_{01} - q_{99}$ support and rescaling it to $[0,1]$ keeps the fitted curve focused on the portion of the data where the sample is dense enough to identify meaningful curvature. This also reduces the risk that the spline is driven by a few extreme observations, which can create unstable turning points or misleading tail behavior. In economic terms, stabilization ensures that the estimated shape reflects the relevant interior tradeoff between liquidity, volatility, and price informativeness rather than extrapolation at the extremes.

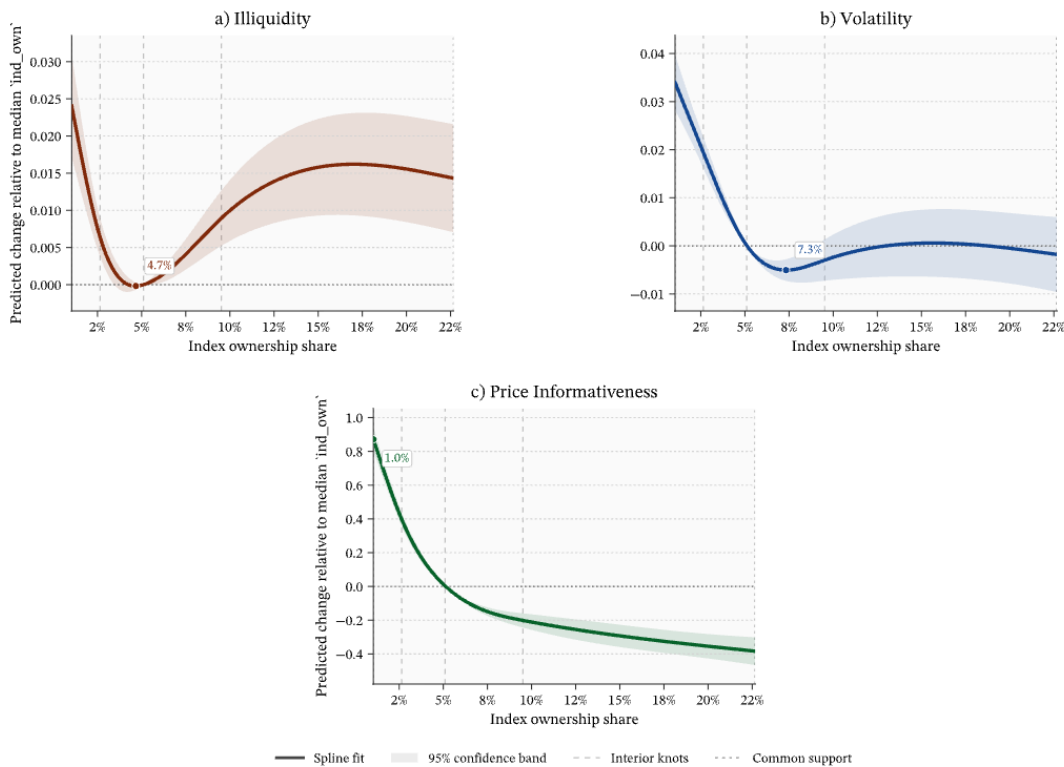


Figure 1.7: Spline Regression Results

*Notes:*Note: This figure plots panel fixed-effects spline estimates of the relationship between index ownership and market quality. Curves are centered at the median index ownership, shaded bands denote point-wise 95% confidence intervals, and vertical dashed lines mark common support and interior knot locations. Labeled points indicate the outcome-specific optimum implied by the spline.

Figure 1.7 summarizes the channel-by-channel fitted curves, showing a clear non-linear relationship between index ownership and market quality. For Amihud illiquidity and volatility, the fitted curves improve as ownership moves out of the lower tail and into the moderate range. This is consistent with the interpretation that passive ownership can reduce trading frictions and stabilize prices at intermediate ownership levels. Price informativeness is weaker and less precisely pinned down, suggesting that the information channel is weaker than the liquidity and volatility channels in this specification. Taken together, the results imply that the effect of indexing is not

constant across the ownership distribution, but instead concentrated in the interior region where market functioning appears to be strongest.

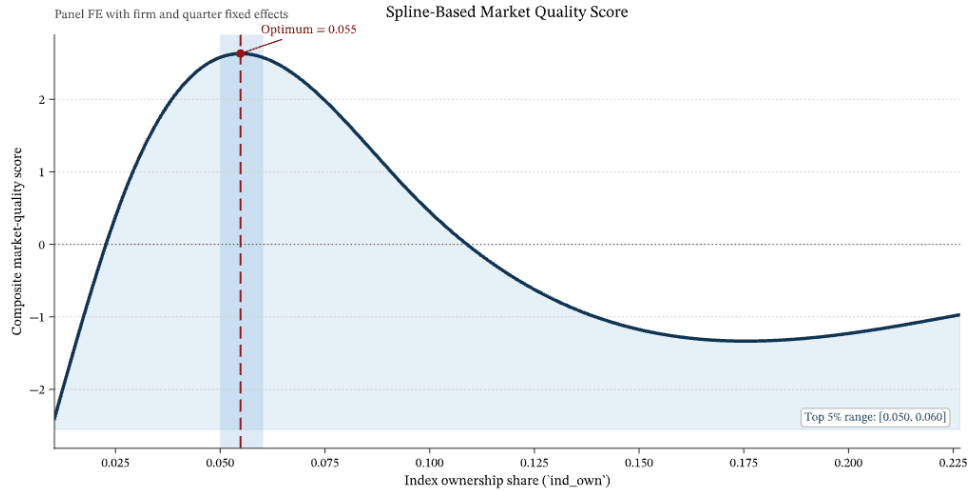


Figure 1.8: Spline Regression Results

*Notes:*Note: This figure plots the composite market-quality score implied by the spline specification. Higher values indicate a more favorable combination of lower illiquidity, lower volatility, and higher price informativeness. The shaded region marks the top 5% of score values, and the vertical marker identifies the implied optimal range of index ownership.

Figure 1.8 traces the composite score across the ownership grid. In the stabilized reduced-form spline, the score peaks at approximately 0.0548, and the top 5% interval lies between 0.0500 and 0.0603. That narrow band indicates that the best market-quality outcomes are concentrated in a relatively tight interior region rather than spread broadly across moderate ownership levels. Economically, this is useful because it translates the separate channel responses into one summary measure of where the balance of liquidity, volatility, and informativeness is most favorable.

The score peak is also informative about how the channels interact. Liquidity and volatility improve most clearly in the moderate range, whereas price informativeness does not contribute as strongly to a sharp interior optimum. As a result, the composite score should be interpreted as an economically weighted compromise across market-quality dimensions rather than as a universal threshold for indexing ownership. This

is the main advantage of the score construction: it turns several related outcomes into a single market-quality summary while preserving the fact that the underlying channels do not move in identical ways.

Relative to the quadratic benchmark, the spline specification is valuable because it relaxes the assumption of one smooth global curvature and allows the data to reveal where the relationship bends most sharply. The estimated optimum moves only slightly upward, from about 0.045 in the quadratic exercise to about 0.055 in the stabilized spline, but the substantive conclusion does not change. In both cases, the evidence points to an interior range of ownership in which market quality is strongest. That stability is important because it indicates that the basic result is not driven by a restrictive polynomial form, even though the spline reveals more detail about how the contribution of indexing differs across the low, middle, and high parts of the support.

The spline comparison also clarifies how to read the shift in the estimated optimum. A quadratic fit imposes a single curvature over the full range, so it can summarize the average direction of the effect but not the local economic mechanism. The spline, by contrast, is better suited to identifying whether moderate indexing improves market quality because of a broad interior plateau or because of a sharper turning region. In that sense, the spline result strengthens the reduced-form argument by showing that the interior optimum persists when the functional-form restriction is relaxed, which is exactly the kind of robustness that matters in empirical finance.

1.5.5 SPLINE IV ANALYSIS

The spline analysis is also extended to an instrumental-variables setting using the global Russell 2000 cutoff design.

First Stage:

$$\begin{aligned}
 IO_{iq} = & \pi_0 + \pi_1 Z_{iq} + \pi_2 (Z_{iq} - \kappa_1)^+ + \pi_3 (Z_{iq} - \kappa_2)^+ + \pi_4 (Z_{iq} - \kappa_3)^+ \\
 & + \gamma X'_{iq} + \alpha_i + \lambda_q + \nu_{it}
 \end{aligned} \tag{1.9}$$

Second Stage:

$$\begin{aligned}
 Y_{iq} = & \beta_1 \widehat{IO}_{iq} + \beta_2 (\widehat{IO}_{iq} - \kappa_1)^+ + \beta_3 (\widehat{IO}_{iq} - \kappa_2)^+ + \beta_4 (\widehat{IO}_{iq} - \kappa_3)^+ \\
 & + \delta X'_{iq} + \alpha_i + \lambda_q + \varepsilon_{iq}
 \end{aligned} \tag{1.10}$$

In this extension, the endogenous spline basis in index ownership is instrumented with interactions of the cutoff indicator and running variable powers up to the fourth order, together with the standard firm-size, dollar-volume, and running-variable controls. The same spline knots are retained, and firm and date fixed effects are absorbed through two-way demeaning. This design is important because it allows the flexible spline shape to be estimated in a causal framework rather than only as a reduced-form correlation.

The first-stage diagnostics are surprisingly strong across all three market-quality specifications, with minimum F-statistics remaining around 15.7 to 15.9 even after the endogenous spline basis is expanded into multiple terms. That is an important result because spline IV designs are typically harder to identify than simple linear specifications. The global Russell 2000 IV explains the ownership basis with enough power to support credible second-stage inference. In practical terms, this means the causal extension is not being pushed forward by a weak instrument or by a single favorable outcome; the relevance condition holds consistently across the market-quality measures. The strong first stage therefore gives the spline IV model a rare combination of flexibility and identification strength, making it a materially more convincing causal check than the local nonlinear IV specifications discussed earlier in the paper.

Table 1.5: First-Stage F Statistics for Spline IV at Global Russell 2000 Cutoff

	Amihud illiquidity	Volatility	Price informativeness
Spline basis 1	53.9018 (p=<1e-16)	53.8301 (p=<1e-16)	53.9410 (p=<1e-16)
Spline basis 2	18.5412 (p=<1e-16)	18.4897 (p=<1e-16)	18.4978 (p=<1e-16)
Spline basis 3	15.7369 (p=1.55e-15)	15.9077 (p=9.99e-16)	15.6617 (p=1.89e-15)
Spline basis 4	56.1608 (p=<1e-16)	56.2035 (p=<1e-16)	56.0468 (p=<1e-16)
Observations	235544	234897	234281
Firms	8627	8605	8614
Dates	99	99	99
Cutoff	2000	2000	2000
Poly degree	4	4	4
Min first-stage F	15.7369	15.9077	15.6617
Controls	Yes	Yes	Yes
Fixed effects	firm + date	firm + date	firm + date
SE type	firm clustered	firm clustered	firm clustered

Notes: This table reports first-stage relevance diagnostics for the spline IV specification. Entries are first-stage F statistics for each endogenous spline basis term, with p-values shown underneath. The model uses firm and date fixed effects and firm-clustered standard errors.

* p<0.10, ** p<0.05, *** p<0.01.

The spline basis is a natural cubic spline in *'ind_own'* with four endogenous terms and excluded instruments formed by cutoff interactions with running-variable powers.

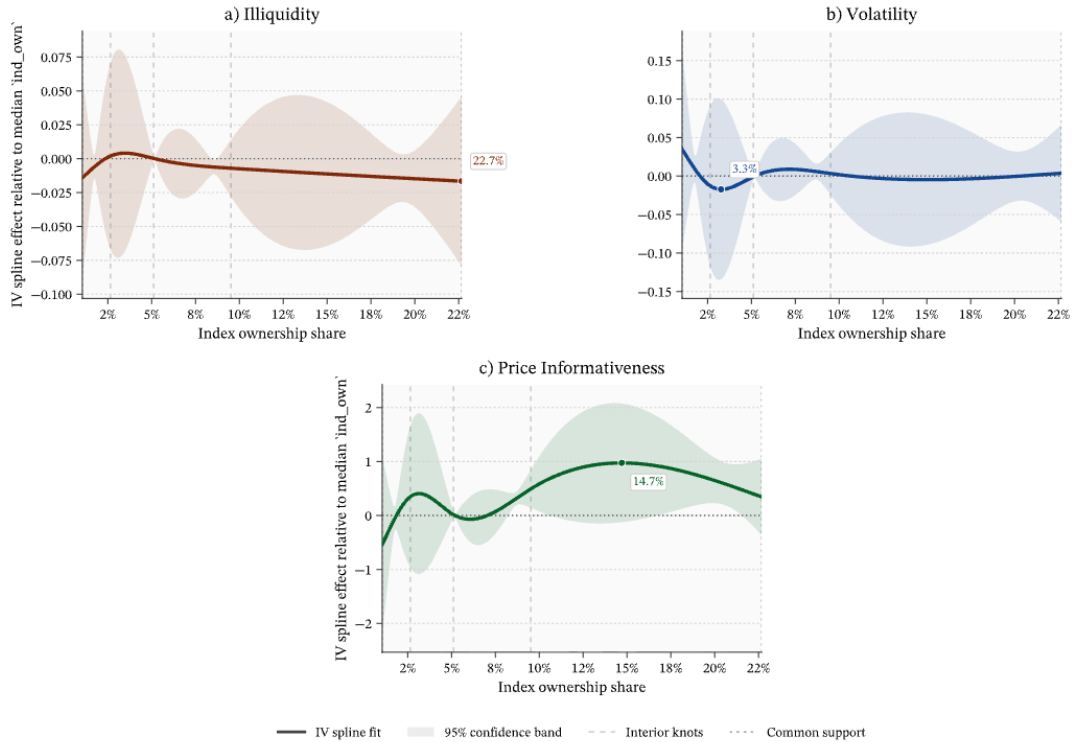


Figure 1.9: Spline Regression Results

*Notes:*Note: This figure shows IV spline estimates of the nonlinear effect of index ownership on each market-quality outcome under the global Russell cutoff 2000 design. Curves are centered at the median `ind_own`, shaded bands denote pointwise 95% confidence intervals, and vertical dashed lines mark spline knot locations. The figure highlights heterogeneous causal nonlinearities across liquidity, volatility, and price informativeness.

Figure 1.9 shows a clear nonlinear causal pattern. The results differ across market quality measures: volatility displays the cleanest curvature, while liquidity and price informativeness bend more unevenly. The fitted line it bends in a way that is consistent with an interior tradeoff between the benefits and limits of passive ownership. At lower ownership levels, additional indexing appears to improve market functioning. This is consistent with the idea that a larger passive presence can reduce trading friction, deepen liquidity provision, and dampen short-run price pressure. As indexing

increases, the curve then flattens and eventually changes slope, which is the standard sign of diminishing marginal gains from additional passive ownership.

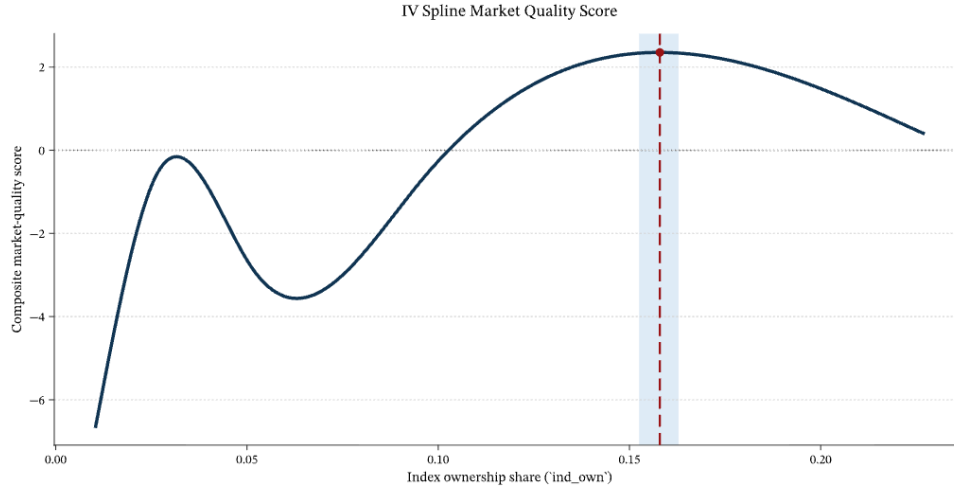


Figure 1.10: Spline Regression Results

*Notes:*Note: This figure reports the composite market-quality score based on the IV spline estimates under the global cutoff 2000 design. The peak of the curve identifies the causal ownership level associated with the strongest overall market-quality outcome, while the shaded band indicates the top 5% range of score values.

Figure 1.10 summarizes those fitted relationships in a single market-quality score. The composite optimum lies at approximately 0.1579, with the top 5% interval between 0.1525 and 0.1628. Visually, this confirms that the best causal market-quality outcome is concentrated in a relatively narrow interior range rather than being spread across the ownership distribution. Another interesting takeaway is that the most favorable ownership region farther to the right than in the reduced-form benchmark. The key implication is that once endogeneity is removed, the strongest joint market-quality outcome appears at a higher ownership level. In other words, the market needs a larger passive presence before the combined liquidity and volatility gains dominate the tradeoff.

Taken together, the reduced-form spline and the causal spline IV extension support the same broad conclusion: indexing ownership is associated with market quality in

a nonlinear way, and the best outcomes occur in an interior region of the ownership distribution. The reduced-form model places that region around 5% to 6% ownership, while the IV extension shifts the composite causal optimum upward toward roughly 15% to 16% ownership. The difference between the two estimates should not be read as a contradiction. Rather, it shows that the shape of the relationship is sensitive to endogeneity correction and the spline IV extension is valuable.

The main value of the spline IV extension is that it strengthens both identification and robustness. On the identification side, the preferred Russell 2000 design delivers strong first-stage relevance for the entire spline basis, so the second stage is strong and meaningful. On the robustness side, the spline relaxes the quadratic restriction and shows that the interior optimum survives when the functional form is allowed to vary more flexibly across the support. This combination matters because it rules out two common concerns at once: that the result is merely a polynomial artifact, and that the causal pattern is unstable once the ownership relation is estimated more carefully.

More broadly, the spline analysis indicates that the relationship between indexing and market quality is nonlinear and that the economically relevant ownership range is interior rather than extreme. The spline analysis complements the quadratic analysis by confirming interior nonlinearity under a more flexible specification and by showing that the causal shape remains nonlinear once the design is extended to the Russell 2000 IV framework. This section strengthens the discussion as it adds credibility to both functional-form robustness and identification.

1.6 CONCLUSION

This dissertation examines how index ownership affects market quality across liquidity, volatility, and price informativeness. The baseline two-way fixed-effects regressions provide a descriptive benchmark: higher index ownership is associated with higher Amihud illiquidity ($\beta = 0.0613$), lower realized volatility ($\beta = -0.0560$), and lower values of the synchronicity-based informativeness proxy ($\beta = -3.3018$). These mixed signs indicate that passive ownership operates through multiple channels rather than a single monotonic mechanism. Accordingly, the baseline results motivate identification and functional-form analysis to distinguish correlation from causal effects and to test for nonlinear responses.

The instrumental-variables evidence uses Russell 1000/2000 reconstitution to isolate exogenous variation in ownership. At the Russell 1000 cutoff, index ownership is associated with a statistically precise reduction in volatility, while effects on illiquidity and informativeness are directionally consistent but imprecise, implying that the volatility channel is the most robust causal margin for large-cap firms. At the Russell 2000 cutoff, the ownership shock is larger and yields a sizable liquidity improvement ($\beta = -0.2849$) alongside a pronounced decline in the informativeness proxy ($\beta = -12.0868$), with volatility effects weaker. This pattern suggests that for smaller firms, passive inflows can deepen trading while simultaneously reducing firm-specific information in prices. Taken together, the IV results indicate that causal effects vary across cutoffs and firm environments, and that liquidity gains are most evident for smaller firms experiencing larger passive inflows.

Nonlinear analysis refines the interpretation of these mixed linear results. Quadratic fixed-effects models reveal U-shaped patterns for volatility and price informativeness, while illiquidity exhibits weaker curvature; a composite market-

quality score peaks at index ownership around 0.0379 with a high-quality range of roughly 3.3%~4.3%. IV-based nonlinear estimates confirm a robust U-shape for volatility at the Russell 2000 cutoff, with a turning point near 0.0809, while curvature in other channels is less stable under causal identification.

The spline analysis provides the strongest evidence that indexing ownership affects market quality in a nonlinear way. In the reduced-form spline specification, the composite market-quality optimum lies in a moderate ownership range of roughly 5% to 6% with the highest score concentrated in a narrow interval around 0.0500 to 0.0603. The spline IV extension strengthens the interpretation the causal interpretation: under the global Russell 2000 identification design, the nonlinear pattern remains robust after addressing endogeneity, preserving an interior optimum while shifting the favorable ownership range upward to approximately 15% to 16%, and maintaining a strong first stage across the full spline basis. Importantly, these results show that the interior optimum is not a quadratic artifact but a stable feature of the data that survives causal identification. Across both quadratic and spline specifications, a consistent conclusion emerges: the most favorable market outcomes occur in an interior ownership range, with volatility providing the clearest and most robust curvature and liquidity improving most clearly at moderate ownership levels.

This analysis has several limitations qualify the interpretation of the results. First, index ownership is measured from index mutual funds and ETFs at the quarterly frequency, so the analysis may understate other passive channels and may smooth short-run liquidity dynamics that occur within quarters. Second, while the Russell reconstitution provides credible identification, the exclusion restriction is strongest near the cutoff and the nonlinear IV design remains sensitive to weak-instrument concerns for squared ownership. Third, the informativeness proxy is based on return synchronicity, which captures comovement but may not map one-to-one to firm-specific information

production. Finally, the sample covers U.S. equities from 1993–2023, so external validity to other markets, asset classes, or regulatory regimes is uncertain.

Despite these limitations, the combined baseline, IV, and nonlinear evidence provides a coherent picture of how passive ownership relates to market quality and identifies the ranges where the trade-offs are most pronounced. Overall, the evidence points to a trade-off between liquidity provision and information production in the presence of passive ownership. Baseline correlations and causal estimates both suggest that indexing can lower volatility and improve liquidity in some regimes, yet may also weaken price informativeness, especially when passive ownership becomes dominant. The nonlinear results reconcile these tensions by showing that the net effect depends on the ownership level and firm environment, rather than a single average coefficient. This conclusion underscores that the market-quality effects of indexing are not uniform and that policy discussions should focus on marginal changes around prevailing ownership levels rather than on extremes.

CHAPTER 2

HETEROGENEOUS EFFECTS OF INDEX OWNERSHIP: EVIDENCE FROM FIRM SIZE AND BUSINESS CYCLE CONDITIONS

2.1 INTRODUCTION

The growth of index ownership has altered the investor base of U.S. equities and, in turn, raised a core question for market quality: when does passive ownership improve trading conditions and when does it weaken information production? The baseline regressions show mixed average effects across liquidity, volatility, and a synchronicity-based informativeness proxy, which indicates that a single average coefficient is unlikely to capture the full pattern. As a result, this chapter develops a structured heterogeneity analysis to map how ownership–market-quality correlations vary across environments. The objective is descriptive rather than structural, and the analysis is designed to provide additional evidence on where the trade-offs are strongest rather than to identify an optimal ownership level.

The heterogeneity analysis proceeds along three dimensions. First, the size-sorted design tests whether ownership effects differ across the firm-size distribution, where information frictions and trading depth vary systematically. Second, the business-cycle design evaluates whether ownership effects shift between expansions and recessions, when market-wide liquidity and risk conditions change sharply. Third, the combined size–cycle design examines joint heterogeneity to determine whether size gradients are

amplified or dampened in downturns. This progression provides a structured mapping of ownership effects across the cross-section and the time series.

A large literature highlights that the effects of index ownership on market quality are heterogeneous across firms, particularly by size. Large-cap firms in major indices often experience improved liquidity and narrower spreads, while small-cap firms may face weaker or adverse effects due to lower trading volumes and market depth (Ben-David, Franzoni, and Moussawi, 2018; Petajisto, 2013; Cremers and Petajisto, 2009). Firm size also shapes trading dynamics, with larger firms attracting more informed and institutional trading that improves price efficiency (Chordia, Roll, and Subrahmanyam, 2000; Hasbrouck, 2007). To capture this heterogeneity, scholars employ size-sorted regressions that estimate ownership effects within deciles or quintiles while controlling for firm and time factors (Brennan and Subrahmanyam, 1996; Foucault, Kadan, and Kandel, 2005; Acharya et al., 2013). This approach provides a granular view of how indexing effects vary across the size distribution.

A related literature emphasizes state dependence over the business cycle. Market liquidity, volatility, and investor risk appetite shift between expansions and recessions, which can amplify or mitigate passive-ownership effects (Foucault, Sraer, and Thesmar, 2010; Ben-David, Franzoni, and Moussawi, 2018; Adrian, Boyarchenko, and Giannone, 2019). Expansions tend to absorb passive flows with limited disruption, whereas recessions or financial stress can exacerbate volatility and bid-ask spreads, especially for smaller or less liquid firms (Bekaert, Harvey, and Lundblad, 2007; Kelly, Pruitt, and Su, 2018). Market microstructure research also shows that liquidity provision and trading dynamics vary with macro conditions (Brunnermeier and Pedersen, 2009; Acharya et al., 2013). These findings underscore the need to model cyclical heterogeneity when evaluating index ownership.

This study contributes to the literature on index ownership and market quality in three ways. First, it extends prior work on firm-size heterogeneity by systematically analyzing how passive ownership affects liquidity, volatility, and price informativeness across size deciles. Second, it incorporates business-cycle dynamics to test whether ownership effects shift between expansions and recessions, highlighting state-dependent market responses. Third, and most importantly, it integrates these two dimensions by examining how the interplay between firm size and macro conditions shapes the impact of indexing on market quality. By jointly considering size and cyclical effects, the analysis provides a more nuanced answer to the question of how much indexing is good for the market and informs both academic debates and policy discussions.

The empirical strategy builds on the baseline two-way fixed-effects framework with controls for firm size and dollar trading volume. For size heterogeneity, the specification is estimated separately within market-capitalization deciles. For cycle heterogeneity, index ownership is interacted with a recession indicator to capture slope shifts between downturns and expansions. For joint heterogeneity, ownership effects are estimated within each size decile and cycle segment, producing a grid of coefficients summarized by heatmaps. All specifications retain firm and time fixed effects and two-way clustered standard errors.

The remainder of the chapter proceeds in this order. Section 2.1 presents size-sorted results and highlights the strongest cross-sectional patterns. Section 2.2 turns to the business-cycle interaction and documents state dependence in liquidity, volatility, and informativeness. Section 2.3 reports the combined size-cycle analysis and interprets joint heterogeneity across outcomes. Section 2.4 concludes with limitations.

2.2 SIZE-SORTED ANALYSIS

2.2.1 SET-UP

Firm size is a core determinant of trading intensity and information environments. Smaller firms face higher information frictions, thinner liquidity, and more segmented investor bases, whereas larger firms trade more frequently and are covered by broader information intermediaries. Consequently, identical changes in passive ownership may trigger different responses in liquidity provision, volatility, and price informativeness across size groups.

The previous chapter documents that index ownership is associated with multiple, and sometimes opposing, dimensions of market quality. In the IV analysis, ownership effects are opposite across the 1000/2000 cutoffs: near the Russell 1000 cutoff (large-cap boundary) the IV estimates indicate a statistically precise decline in volatility, whereas near the Russell 2000 cutoff (small-cap boundary) the IV estimates indicate liquidity improvements alongside a pronounced decline in the informativeness proxy. This evidence motivates a systematic heterogeneity analysis that conditions on firm size to identify more precisely where the impact of index ownership changes.

This section evaluates heterogeneity in the relationship between index ownership and market quality across firm size. The heterogeneous analysis uses size-sorted regressions to map the ownership–market-quality relationship across the full distribution of market capitalization. It provides an average association, while the size-sorted design asks whether the same mechanisms operate for small and large firms with distinct information environments and trading frictions. The analysis sorts observations into market-capitalization deciles and estimates identical fixed-effects regressions within each decile. This approach isolates cross-sectional heterogeneity while holding the

specification constant, so any variation in coefficients can be attributed to size-related differences.

The empirical design mirrors the baseline specification while allowing the coefficient on index ownership to vary by firm size.

$$Y_{i,q} = \alpha_i + \lambda_q + \sum_{k=1}^{10} \beta_k (\text{IndexOwn}_{i,q} \times D_k) + \gamma X'_{i,q} + \epsilon_{i,q} \quad (2.1)$$

Variable construction follows the baseline chapter. Amihud illiquidity is computed from daily returns and dollar volume and aggregated to the quarter; volatility is the quarterly standard deviation of daily returns; and price informativeness measure is the synchronicity-based price-informativeness proxy derived from a factor-model R^2 transformation. Index ownership uses quarter-end holdings of index mutual funds and ETFs relative to shares outstanding. The same data sources and filters apply, so size-based comparisons reflect heterogeneity rather than measurement changes.

The sample is partitioned into ten market-capitalization deciles and each decile contains approximately 23,928 to 23,929 observations. Sorting is performed at the observation level using market capitalization, ensuring that the within-decile regressions compare firms of similar size at each date. The regressions use the ownership-covered subsample with consistent coverage for the three outcomes. Because the same specification is estimated within each decile, differences in coefficient patterns are not driven by model variation or uneven sample size. At the same time, these estimates remain correlational and should be interpreted as descriptive heterogeneity rather than causal effects.

2.2.2 SIZE EFFECTS

The size gradient is most pronounced for the price informativeness proxy. Index ownership is negatively associated with the proxy in deciles 1 through 9, with large t-

statistics in the small-cap tail (for example, $\beta = -13.60$ and $t = -10.47$ in decile 1, and $\beta = -8.21$ and $t = -12.47$ in decile 2). The magnitude attenuates steadily as size increases, and the coefficient becomes small and statistically insignificant in the largest decile ($\beta = -0.40$, $t = -0.50$). This pattern indicates that the negative association between passive ownership and the informativeness proxy is concentrated among smaller and mid-sized firms rather than among the largest firms. Figure 2.1 summarizes the size-sorted coefficients and their confidence intervals.

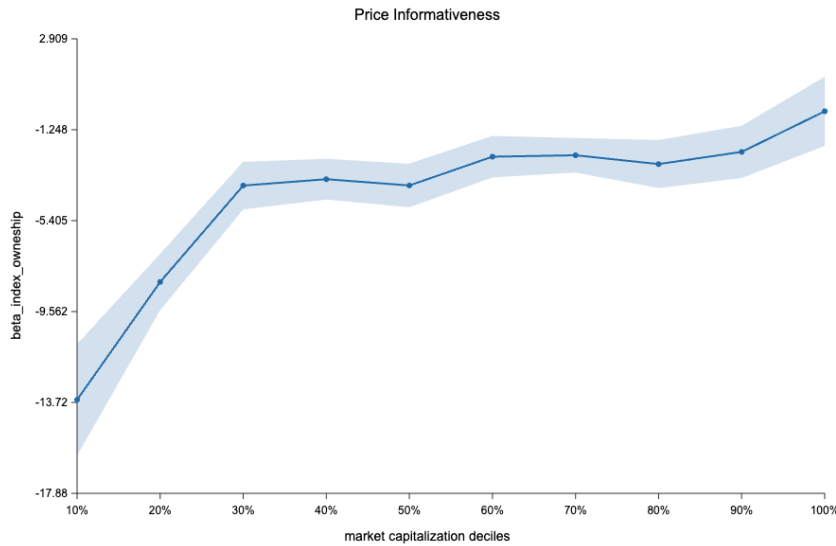


Figure 2.1: Size-sorted Regression Results

*Notes:*Note: Size-sorted coefficient estimates for index ownership on the price informativeness proxy with 95% confidence intervals..

The economic interpretation is consistent with the mechanism highlighted in the prior chapter: passive ownership can reduce incentives for firm-specific information production, and this effect appears strongest where baseline information frictions are high. Because small firms rely more heavily on marginal information producers, a shift toward passive ownership may disproportionately reduce the incorporation of firm-specific signals into prices. By contrast, for large firms with deep analyst coverage and abundant public information, the same ownership changes appear less

consequential for the informativeness proxy. This size pattern therefore reconciles the baseline negative coefficient with the IV evidence at the Russell 2000 cutoff, where the informativeness decline is most pronounced. At the same time, the findings underscore that the informativeness channel is not uniform across the cross-section.

Liquidity effects, measured by the Amihud illiquidity proxy, are comparatively weak across size groups. Only the third decile exhibits a statistically significant estimate ($\beta = 0.019$ and $t = 2.28$), while the remaining deciles yield economically small and statistically indistinguishable coefficients. The coefficient range from -0.068 to 0.019 suggests that any size-related variation in the illiquidity response is modest relative to the variation observed in the informativeness proxy. This lack of a clear monotone pattern contrasts with the IV evidence near the Russell 2000 cutoff, where liquidity improvements were sizable, and indicates that such effects may be localized to specific ownership shocks rather than broad size categories. Figure 2.2 reports the decile estimates with confidence intervals.

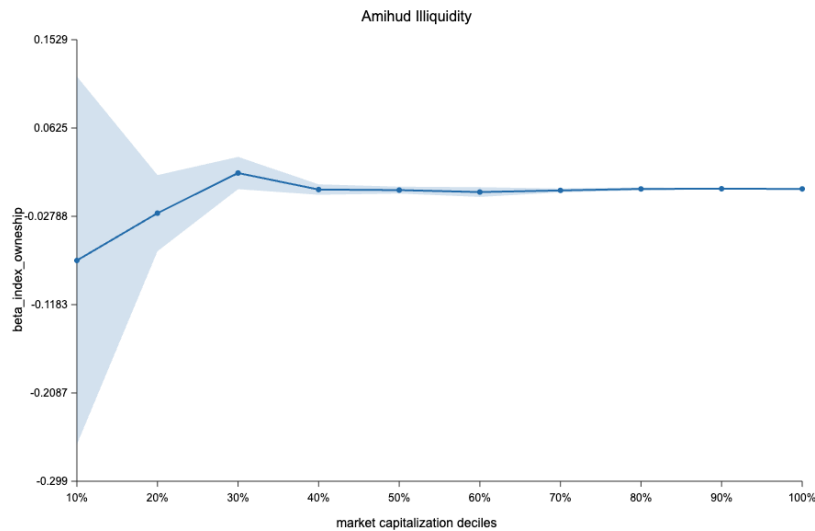


Figure 2.2: Size-sorted Regression Results

*Notes:*Note: Size-sorted coefficient estimates for index ownership on Amihud illiquidity with 95% confidence intervals..

One interpretation is that liquidity changes induced by passive ownership are mediated by trading demand rather than firm size per se. If passive flows are concentrated around index events or concentrated among specific investor segments, the liquidity response may depend more on event intensity than on cross-sectional size. Alternatively, the Amihud proxy may be less sensitive to small changes in ownership within a size group once firm and date fixed effects are included. In either case, the size-sorted design suggests that liquidity is the least systematic channel in the cross-section relative to informativeness and volatility. This conclusion highlights the value of complementing size-sorted regressions with event-based identification for liquidity effects.

Volatility exhibits a clearer size dependence than illiquidity, with statistically significant effects concentrated in the middle deciles. Deciles 4 through 6 show negative and significant coefficients (β between -0.139 and -0.106, with t-statistics around -2.1 to -2.3), indicating lower idiosyncratic volatility as index ownership rises for mid-cap firms. The smallest decile has a similar sign but only marginal significance ($t = -1.93$), while the largest deciles show small and statistically insignificant positive coefficients. This pattern suggests that the volatility channel is most active where trading is sufficiently deep to transmit ownership changes, yet information production is not as fully saturated as in the largest firms. Figure 2.3 illustrates the mid-cap concentration of the statistically significant effects.

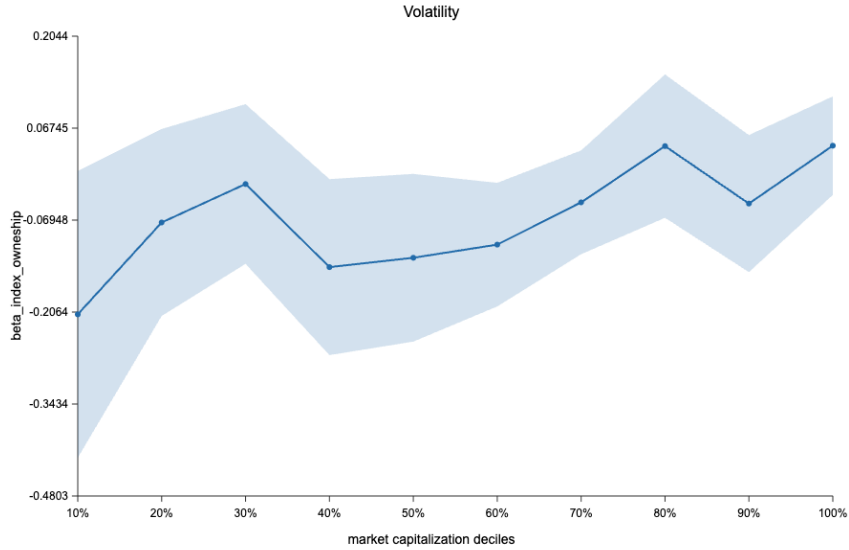


Figure 2.3: Size-sorted Regression Results

*Notes:*Note: Size-sorted coefficient estimates for index ownership on idiosyncratic volatility with 95% confidence intervals..

The mid-cap emphasis also aligns with the prior chapter’s IV evidence at the Russell 1000 cutoff, which identified volatility reductions for large-cap firms but weaker effects at other cutoffs. In the size-sorted results, the volatility effect is not monotone with size, implying that the ownership–volatility relationship depends on a balance between liquidity depth and information sensitivity. This interpretation is consistent with a view in which passive ownership dampens idiosyncratic price movements by increasing common ownership and stabilizing flows, but only within a particular range of firm size. Outside that range, either trading is too thin for flow-based stabilization or information production is too robust for ownership shifts to materially alter volatility. The nonlinearity across size reinforces the broader conclusion that ownership effects are context dependent.

The size-sorted evidence indicates that the informativeness channel is the most consistently heterogeneous dimension of market quality. It suggests that size interacts more strongly with information channels than with trading-cost channels. Small firms

typically face thinner analyst coverage and higher idiosyncratic information frictions, so index ownership may affect the price-information proxy more strongly in those segments. Liquidity in large firms is already high and may be less sensitive to changes in index-fund ownership, which is consistent with flat illiquidity coefficients. The volatility pattern aligns with diminishing marginal effects, where incremental index ownership may stabilize returns when baseline volatility is high but becomes negligible for large firms. In summary, information production costs dominate for smaller firms, whereas volatility stabilization is more relevant for mid-cap firms.

While the size-sorted analysis is informative, it potentially misses another important factor: cyclical effects. The influence of index ownership on market conditions may vary with broader macroeconomic and financial environments, and market quality can shift sharply across phases of the business cycle. Liquidity and volatility are particularly cyclical, which implies that the same ownership change may have different implications in expansions than in recessions. As a result, a business-cycle-based heterogeneous analysis is conducted to examine how the ownership–market-quality relationship differs across expansion and recession regimes.

2.3 BUSINESS CYCLE ANALYSIS

2.3.1 SET-UP

Business-cycle conditions are a central determinant of market quality, and they provide a natural dimension of heterogeneity for the effects of index ownership. Liquidity, volatility, and the informational content of prices move with macroeconomic conditions, and these shifts can alter how passive ownership translates into trading outcomes. The time-series summaries from the data description section in previous chapter indicate that market-quality proxies and index ownership display pronounced

variation across the sample, which motivates an explicit interaction between ownership and recession states. Accordingly, this section asks whether the association between index ownership and market quality differs in recession versus expansion regimes. By conditioning on business-cycle states, the analysis complements the size-sorted results and clarifies whether ownership effects are amplified or muted when market conditions deteriorate.

The recession indicator follows NBER-style dating and covers 2001Q1–2001Q4, 2007Q4–2009Q2, and 2019Q4–2020Q2. These episodes coincide with periods of heightened macroeconomic stress and substantial dislocations in trading activity. In the time-series summaries, the ownership measure and market-quality proxies display clear movements over time, reinforcing the need to allow ownership effects to differ across these regimes. This background motivates a panel specification that interacts index ownership with a recession dummy while retaining firm and time fixed effects. The goal is to test whether the ownership–market-quality relationship shifts in economically meaningful ways during downturns.

The business-cycle heterogeneity specification follows the baseline two-way fixed-effects design but adds an interaction between index ownership and a recession indicator.

$$Y_{iq} = \alpha_i + \lambda_q + \beta IO_{iq} + \delta (IO_{iq} \times \text{Recession}_q) + \gamma X'_{iq} + \varepsilon_{iq} \quad (2.2)$$

The regression is estimated separately for Amihud illiquidity, idiosyncratic volatility, and the synchronicity-based price informativeness proxy. The sample contains approximately 239,758–241,199 firm-quarter observations with about 8,855–8,877 firms and 100 time periods, depending on the outcome. Because the model includes time fixed effects, the main recession dummy is absorbed for identification of the interaction, and the coefficients should be interpreted as within-firm slope shifts between

recessions and expansions. The results section reports the expansion-state slopes and the incremental recession adjustments.

2.3.2 CYCLIC EFFECTS

The liquidity channel exhibits the strongest business-cycle heterogeneity. In expansions, higher index ownership is associated with higher Amihud illiquidity ($\beta = 0.080$ and $t = 3.37$), indicating weaker liquidity when passive ownership rises. During recessions, the interaction term is negative and statistically significant ($\delta = -0.2142$, $t = -2.47$), implying that the ownership–illiquidity slope is substantially lower in downturns. Taken together, the recession-state slope equals the sum of β and δ , which implies a materially different relationship between ownership and liquidity in recessions than in expansions. Figure 2.4 summarizes the coefficient estimates and their confidence intervals.

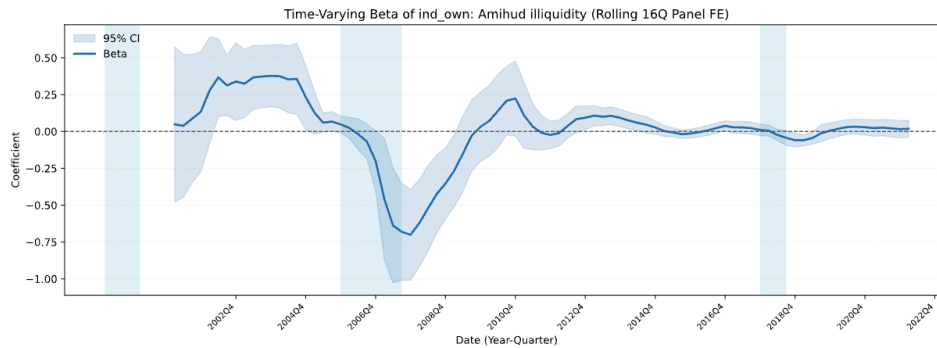


Figure 2.4: Business-cycle Regression Results

*Notes:*Note: Business-cycle interaction estimates for index ownership on Amihud illiquidity with 95% confidence intervals.

A plausible interpretation is that passive ownership interacts with macro liquidity conditions. When trading is strained in recessions, index ownership may provide a stabilizing base of demand or facilitate rebalancing activity that offsets adverse liquidity pressures. By contrast, in expansions the same ownership changes may coincide with

a gradual shift away from active trading that increases illiquidity on the margin. This recession-specific attenuation is consistent with the notion that the liquidity channel is state dependent and should not be summarized by a single average coefficient. The next sections examine whether similar state dependence appears in volatility and informativeness.

For idiosyncratic volatility, the expansion-state coefficient is negative and statistically significant ($\beta = -0.0681$ and $t = -2.95$), indicating lower volatility as index ownership rises in normal times. The recession interaction is positive but statistically imprecise ($\delta = 0.1390$, $t = 1.34$), implying that the volatility-reduction effect weakens in downturns. While the interaction is not significant, the sign suggests that the stabilizing effect of passive ownership is less reliable when macro conditions deteriorate. This pattern aligns with the broader thesis that ownership effects depend on the surrounding market environment rather than being constant over time. Figure 2.5 reports the interaction estimates for volatility.

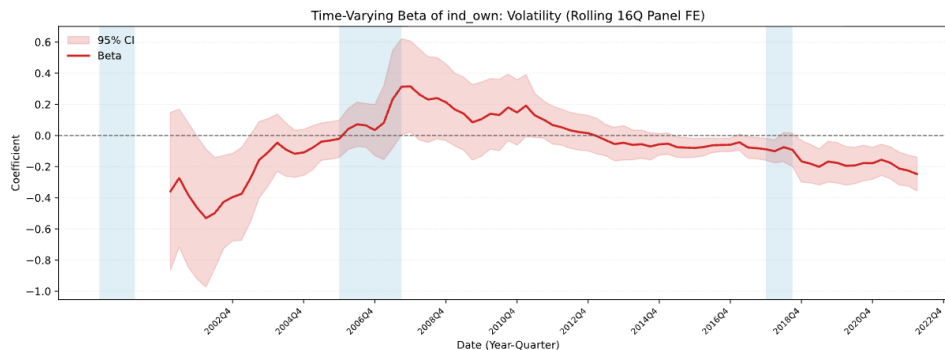


Figure 2.5: Business-cycle Regression Results

*Notes:*Note: Business-cycle interaction estimates for index ownership on idiosyncratic volatility with 95% confidence intervals.

The attenuation of volatility effects in recessions is economically intuitive. Periods of stress often involve large common shocks, correlated trading, and elevated risk premia that dominate firm-specific ownership changes. Under such conditions, the

ability of passive ownership to damp idiosyncratic volatility may be limited by broader market forces. Conversely, in expansions when common shocks are smaller, ownership shifts may more effectively reduce firm-level volatility. This explanation highlights why the volatility channel appears more state dependent than the baseline average suggests.

The price informativeness proxy shows a negative association with index ownership in expansions ($\beta = -3.2530, t = -13.55$), consistent with the baseline evidence that higher passive ownership coincides with less informative prices. The recession interaction term is also negative but statistically insignificant ($\delta = -0.5590, t = -0.70$), implying that the informativeness decline does not materially differ between recessions and expansions. This result suggests that the information-production channel operates persistently across macro regimes, rather than being concentrated in downturns. Figure 2.6 summarizes the interaction estimates for the informativeness proxy.

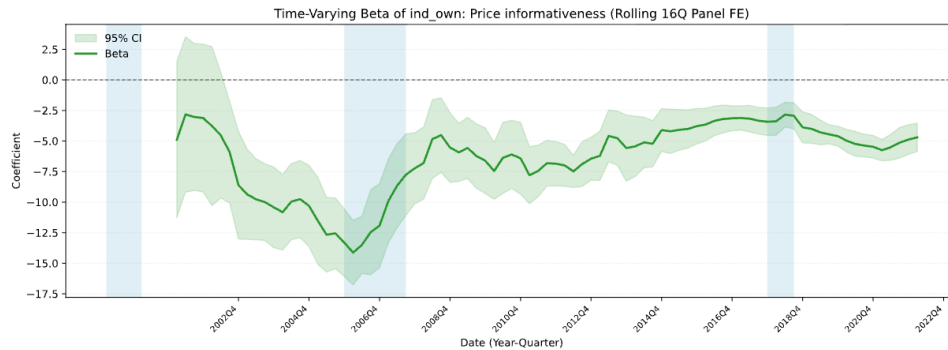


Figure 2.6: Business-cycle Regression Results

*Notes:*Note: Business-cycle interaction estimates for index ownership on the price informativeness proxy with 95% confidence intervals.

One interpretation is that information production responds to ownership structure more than to cyclical conditions. If passive ownership reduces incentives for firm-specific information acquisition in a persistent way, then recession conditions may not further amplify the effect. Alternatively, measurement limitations of the synchronicity

proxy may attenuate state dependence, since macro shocks already raise comovement across firms. These considerations indicate that the informativeness channel is robust but relatively insensitive to cyclical regimes, in contrast to the liquidity channel. This distinction is important for interpreting the overall market-quality trade-off over time.

The business-cycle heterogeneity results reveal that the liquidity channel is the most state dependent, while the informativeness channel is more stable across regimes. In expansions, higher index ownership coincides with higher illiquidity and lower volatility, whereas in recessions the ownership–illiquidity slope is substantially lower and the volatility effect weakens. The informativeness proxy remains negatively associated with ownership in both regimes, suggesting a persistent information–production trade-off. Together, these findings imply that the market-quality consequences of passive ownership depend not only on firm characteristics but also on macroeconomic conditions. This time-variation complements the size-sorted results and indicates that policy discussions should be conditioned on prevailing market states.

The business-cycle analysis reveals systematic variation in ownership effects across macro regimes, while the size-sorted analysis in the previous section reveals cross-sectional differences across firms. Taken together, these two dimensions of heterogeneity may provide a descriptive answer to the question: how much indexing is good for the market. This answer does not imply a structural optimum, but it motivates a heterogeneous analysis that conditions simultaneously on both firm size and business-cycle state. This design would test whether the recession–expansion slope shifts are concentrated among small firms, mid-cap firms, or large firms, and whether size gradients are amplified or dampened in downturns.

2.4 SIZE AND BUSINESS-CYCLE INTERACTIONS

2.4.1 INTERACTION SET-UP

Size-sorted results indicate that the informativeness channel is strongest among small and mid-sized firms, while the business-cycle analysis shows that liquidity and volatility responses shift across expansions and recessions. This section therefore combines both dimensions and examines whether ownership effects differ jointly by firm size and by cycle regime. The exercise is descriptive rather than structural; its goal is to provide additional evidence on when indexing appears most associated with better or worse market conditions.

This joint heterogeneity design is motivated by the idea that the impact of passive ownership is not uniform across either firms or macro states. Smaller firms have higher information frictions and thinner trading, while recessions bring broad liquidity and risk shocks that can overwhelm firm-level dynamics. If the ownership–market-quality relationship is state dependent, then averaging across both dimensions may obscure the environments where the trade-offs are most pronounced. Thus, the combined analysis addresses a descriptive version of the question, “how much indexing is good for the market,” by identifying where the correlations are strongest and where they attenuate. The analysis does not claim an optimal structural level of indexing but instead maps ownership effects across size and cycle regimes.

The empirical strategy extends the baseline two-way fixed-effects regression by interacting index ownership with both firm-size deciles and business-cycle regimes. For each outcome, the analysis estimates separate slopes for each size decile within each cycle segment, using the same controls for firm size and dollar trading volume and retaining firm and time fixed effects. Business-cycle segments follow the

NBER-style dating used earlier and include the recession intervals 2001Q1–2001Q4, 2007Q4–2009Q2, and 2019Q4–2020Q2, as well as surrounding expansion windows.

$$Y_{iq} = \sum_{d=1}^{10} \sum_{b \in \{0,1\}} \beta_{db} (\text{IndexOwn}_{iq} \times \mathbf{1}\{\text{SizeDecile}_i = d\} \times \mathbf{1}\{BC_t = b\}) + \delta_i + \gamma_q + \mathbf{X}'_{iq} \theta + \epsilon_{iq} \quad (2.3)$$

where

$\mathbf{1}\{\text{SizeDecile}_i = d\}$ = indicator for firm i belonging to size decile d ,

$\mathbf{1}\{BC_t = b\}$ = business-cycle regime (e.g., expansion: $b=0$, recession: $b=1$)

The resulting grid of coefficients can be summarized as a heatmap of t -statistics and by highlighting the most precisely estimated cells within each outcome. This approach emphasizes where ownership effects are most reliably estimated, while still acknowledging that many size–cycle cells have small samples, especially in short recession windows. Because the design remains descriptive, the coefficients should be interpreted as conditional correlations rather than causal effects. The next sections report the main patterns for liquidity, volatility, and price informativeness.

2.4.2 SIZE- AND CYCLE-DEPENDENT EFFECTS

Liquidity effects exhibit meaningful variation across both size and cycle regimes. The most precisely estimated liquidity cell occurs in the 2002Q1–2007Q3 expansion for size decile 7, where index ownership is associated with higher Amihud illiquidity ($\beta = 0.0348$, $t = 3.58$). In the same expansion window, several mid-to-large deciles also show positive and statistically significant estimates, whereas the largest decile displays a negative estimate. This pattern suggests that during mid-2000s expansions, higher ownership was correlated with weaker liquidity in a range of mid-sized firms

but not uniformly across the distribution. Figure 2.7 visualizes the size–cycle liquidity pattern using the coefficient heatmap.

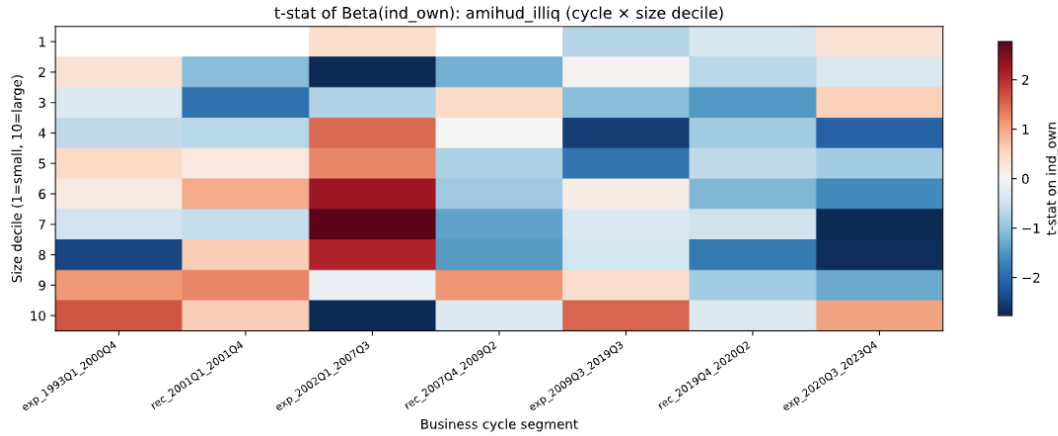


Figure 2.7: Size and Business-Cycle Interactions Results

*Notes:*Note: Heatmap of size-by-cycle coefficients for index ownership on Amihud illiquidity.

During recession windows, liquidity estimates are generally less precise because of shorter samples, and the signs vary by decile. The lack of uniformity across recession episodes indicates that liquidity effects are highly state dependent and may reflect a mixture of passive rebalancing and market-wide stress. This heterogeneity reinforces the earlier conclusion that liquidity is the most regime sensitive channel and that its relationship with ownership should not be summarized by a single average coefficient.

Volatility effects also vary across the joint size–cycle grid. The strongest cell appears in the 2001 recession for size decile 7, where ownership is associated with lower idiosyncratic volatility ($\beta = -1.8811$, $t = -4.46$). This pattern indicates that in early recession conditions, passive ownership coincided with a pronounced reduction in volatility for mid-sized firms. By contrast, many expansion-period cells show small or insignificant coefficients, suggesting that volatility responses are more concentrated in specific regime–size combinations. Figure 2.8 summarizes these results across all size–cycle cells.

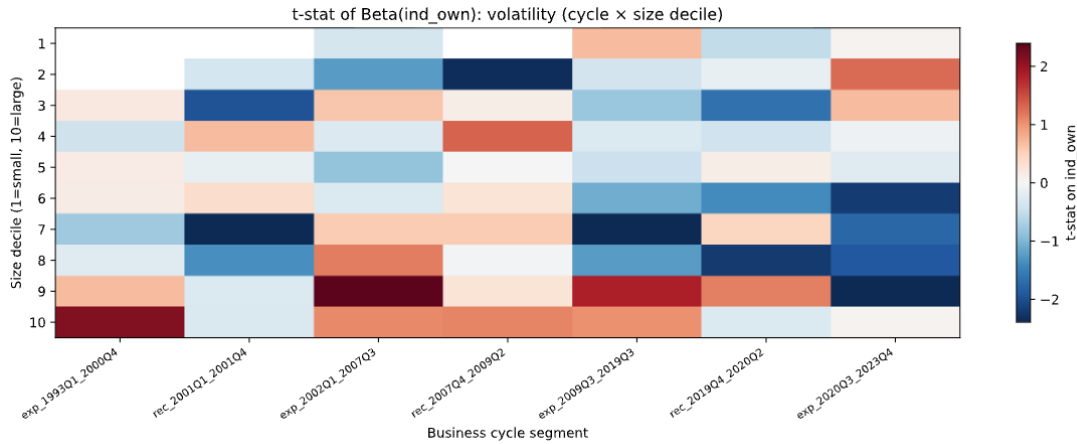


Figure 2.8: Size and Business-Cycle Interactions Results

*Notes:*Note: Heatmap of size-by-cycle coefficients for index ownership on idiosyncratic volatility.

The concentration of significant effects in select recession windows is consistent with the view that volatility responses depend on both firm size and macro stress. When aggregate shocks are large, ownership structure may affect how strongly firm-level volatility is dampened by passive flows, and this effect appears most visible among mid-sized firms. However, the lack of pervasive significance across the grid underscores that the volatility channel is episodic rather than universal. This point cautions against extrapolating any single recession estimate to all cycles or all firm sizes.

The price informativeness proxy displays the most robust and persistent negative association with index ownership across the size-cycle grid. The strongest cell occurs during the 2009Q3–2019Q3 expansion for size decile 4, where the coefficient is large and negative ($\beta = -18.0537$, $t = -9.04$). Across many expansion windows, mid-sized and large firms show consistently negative and significant coefficients, indicating that the informativeness channel remains a central trade-off in periods of normal growth. In recession windows, coefficients are generally less precise but do not overturn the

negative association observed in expansions. Figure 2.9 reports the full heatmap of informativeness estimates.

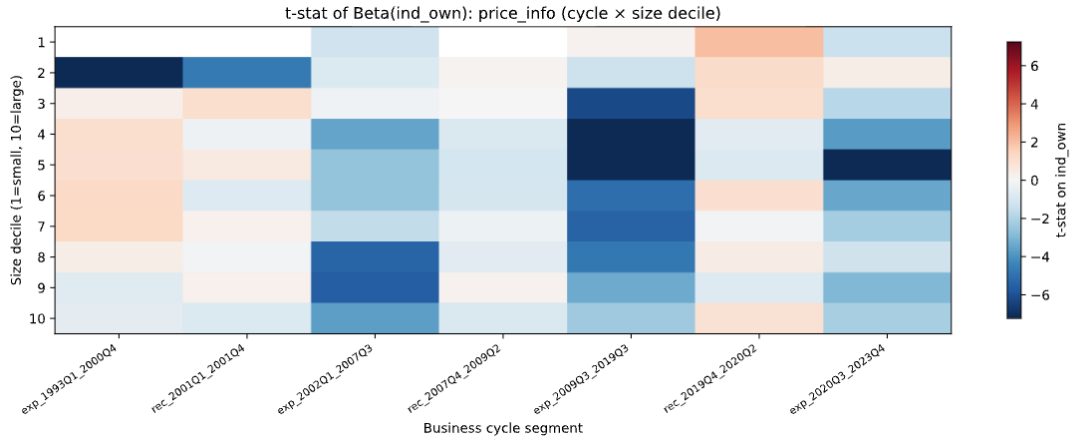


Figure 2.9: Size and Business-Cycle Interactions Results

*Notes:*Note: Heatmap of size-by-cycle coefficients for index ownership on the price informativeness proxy.

The persistence of negative coefficients across size and cycle states suggests that the information-production channel is less cyclical than liquidity and volatility. This finding aligns with the earlier business-cycle interaction results, where the ownership-informativeness slope did not materially differ between recessions and expansions. Therefore, the joint heterogeneity analysis strengthens the case that the informativeness trade-off is a stable feature of passive ownership rather than a purely cyclical phenomenon. At the same time, the magnitude of the effect varies across size deciles, reaffirming that firm characteristics shape the strength of the ownership-informativeness relationship.

The combined heterogeneity results indicate that the market-quality effects of indexing depend jointly on firm size and macroeconomic conditions. Liquidity effects are strongest in specific expansion windows and for mid-sized firms, volatility effects appear episodic in certain recession periods, and informativeness effects are persistently negative across many size-cycle cells. This pattern provides a descriptive answer to

the question of “how much indexing is good for the market”: the apparent benefits and costs are not uniform but depend on the joint environment in which ownership changes occur. Rather than implying a single optimal ownership level, the evidence suggests a range of ownership effects whose sign and magnitude shift with size and cycle. These descriptive results complement the thesis’s broader identification agenda by clarifying where the trade-offs are most visible.

The joint design also offers a framework for connecting the cross-sectional and time-series evidence. If liquidity improvements are concentrated in certain size–cycle cells, then event-based or IV estimates should be interpreted relative to the macro regime and size composition of the affected firms. Similarly, persistent informativeness declines across expansions imply that ownership changes may carry information costs even in benign environments.

2.5 CONCLUSION

This chapter provides a structured heterogeneity analysis of how index ownership relates to market quality. The size-sorted results show that informativeness declines are concentrated among small and mid-sized firms, while liquidity and volatility effects are weaker and less uniform. The business-cycle analysis indicates that liquidity and volatility responses shift between expansions and recessions, whereas the informativeness trade-off persists. The joint size–cycle analysis confirms that the strongest effects concentrate in specific regime–size cells rather than in a single average coefficient.

The joint size–cycle analysis integrates these two dimensions and reveals where the strongest effects concentrate. Liquidity effects are most precisely estimated in specific expansion windows and mid-sized firms, volatility effects are episodic and concentrated in select recession–size cells, and informativeness effects remain broadly

negative across a wide range of size–cycle environments. This combined evidence provides a descriptive answer to the question of how much indexing is good for the market: the apparent benefits and costs depend on the joint firm and macro environment rather than on a single average coefficient. The chapter therefore reframes the debate from a uniform effect to a context-dependent trade-off.

However, there are 3 key limitations. The results remain descriptive and therefore cannot establish causal mechanisms. Second, some size–cycle cells are thin, especially in short recession windows, which reduces precision and increases sensitivity to sampling variation. The business-cycle indicator is binary and may mask variation within expansions or within downturns. Future work should combine the joint heterogeneity framework with quasi-experimental ownership shocks and continuous financial-condition measures to assess whether these patterns persist under causal identification.

Despite these limitations, the heterogenous analysis is informative. The evidence implies that market-quality assessments should account for both firm composition and macro conditions. Information-production incentives appear most sensitive to passive ownership among smaller firms, while liquidity and volatility responses are more cyclical. This context dependence suggests that policy discussions should focus on marginal ownership changes in specific environments rather than on an overall average effect.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis cursus. Duis consectetur, lacus id vehicula viverra, dolor quam dignissim pede, nec ultrices metus tortor quis turpis. Praesent non orci non est tincidunt mollis. Nullam pellentesque auctor orci. Morbi aliquam. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus.

In auctor ornare risus. Nulla id magna nec pede elementum rhoncus. Suspendisse ultricies mattis lacus. Cras pulvinar pede nec dolor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Suspendisse potenti. Sed pede massa, scelerisque id, tempor ac, vestibulum nec, est. Maecenas molestie tincidunt orci. Donec ac leo quis libero iaculis semper. Aenean ligula pede, semper eu, tristique eu, aliquam in, nisi. Curabitur sagittis condimentum ligula. Nunc interdum suscipit libero.

APPENDIX A

FIRST APPENDIX

A.1 RUSSELL RANK CONSTRUCTION

This appendix section documents the construction of the Russell market capitalization ranking used to assign firms to the Russell 1000 and Russell 2000 indices. The procedure follows the replication approach in Ben-David, Franzoni, and Moussawi (2018), which mirrors the annual reconstitution methodology used by FTSE Russell. The objective is to recover a consistent rank ordering when official rankings are not publicly available for the full sample period. The resulting ranks provide a transparent mapping from market capitalization to index assignment, which is central to the empirical design.

Russell indices are reconstituted annually at the end of June, and index membership is determined by market capitalization measured on the rank date, which typically occurs in late May. Firms are ordered by their market capitalization on the rank date, and the largest 1,000 firms are assigned to the Russell 1000, while the next 2,000 firms are assigned to the Russell 2000. This annual sorting produces a sharp change in index weights at the cutoff between ranks 1,000 and 1,001. As a result, firms near the cutoff are similar in size but face materially different index ownership due to benchmark tracking. The ranking universe is restricted to U.S. common stocks with ordinary share codes and primary exchange listings consistent with the standard

CRSP universe. Firms with non-common share codes, non-primary listings, or missing share and price data on the rank date are excluded to ensure that the ranking reflects the eligible investable set. When a firm has multiple share classes, the ranking is based on the primary class observed in CRSP so that market capitalization is not double-counted across classes. These screens align the replicated universe with the investable universe implied by the index methodology and reduce mechanical noise in the rank ordering.

Market capitalization is computed from CRSP daily data using the standard measure of price times shares outstanding. Specifically, for firm i in year t , market capitalization equals the absolute value of price times shares outstanding on the rank date. Because the official rank date is in late May, the last trading day of May is used to approximate the ranking input. This alignment ensures that the rank ordering reflects the information set used in index assignment while preserving a consistent data definition across years. The absolute value of price is used to accommodate CRSP's sign convention for bid-ask midpoint pricing, and shares outstanding are measured in thousands and scaled accordingly. For firms with missing data on the last trading day of May, the most recent prior trading day within May is used to avoid losing observations due to short-lived data gaps. Market capitalization is computed on an unadjusted basis to mirror index construction and to preserve the correct relative ranking across firms. These choices follow common practice in the Russell replication literature and minimize divergence from the official ranking input.

All eligible U.S. common stocks are sorted in descending order by market capitalization on the rank date. The ranking assigns rank 1 to the largest firm and proceeds sequentially through the eligible universe. Index assignment follows the published thresholds: ranks 1 through 1,000 are mapped to the Russell 1000 and ranks 1,001 through 3,000 are mapped to the Russell 2000. This procedure yields a deterministic

index classification that can be reproduced year by year from the underlying CRSP inputs.

Ties in market capitalization are rare at the upper tail, but when they occur the tie is resolved by the CRSP market equity ordering, which is equivalent to the raw ranking by market capitalization. The procedure uses the full ranked list to preserve relative ordering even when firms fall outside the top 3,000, ensuring that the cutoff firms are ranked within a consistent global ordering. Firms entering or exiting the universe between May and June are assigned based on the May rank-date information, consistent with the index methodology that fixes membership at reconstitution. This timing convention maintains a clean separation between ranking inputs and subsequent index membership outcomes.

Because official Russell rankings are not public for the full sample, this study replicates the ranking using CRSP market capitalization data. Following Ben-David, Franzoni, and Moussawi (2018), the replicated ranks are constructed from the rank-date market capitalization and then mapped to the Russell 1000 and Russell 2000 cutoffs. The replicated rankings closely match the index assignments reported in Russell membership files, and the approach is widely used in the literature on index inclusion. Accordingly, the replicated cutoff provides a credible and transparent basis for analyzing discontinuities in index ownership around the 1,000/1,001 threshold.

A.2 IV ROBUTNESS CHECK - FULL RESULTS

This appendix section evaluates the robustness of the IV results to alternative bandwidth choices around the Russell reconstitution cutoffs. The purpose is to test whether the local-window IV estimates confirm the global results or instead reveal weak-instrument behavior. Because identification relies on a discrete ownership jump at

the cutoff, narrowing the window should improve comparability but reduce first-stage variation. The analysis therefore focuses on how relevance, precision, and coefficient stability change across bandwidths. These diagnostics are reported for both the 1,000 and 2,000 cutoffs to assess sensitivity to the source of variation.

The robustness grid sweeps bandwidths of 50, 100, 150, 200, 300, and 500 ranks, in addition to the global sample, for each cutoff. For each window, the same 2SLS specification is re-estimated and diagnostics are recorded, including the first-stage F statistic and the index-ownership coefficient. The outputs include a numerical summary table and two plots that track first-stage strength and coefficient stability across bandwidths. All estimations use the same controls and fixed effects as the baseline IV design to ensure comparability. This grid-based design makes it possible to visualize how identification degrades as the window narrows.

First-stage strength deteriorates sharply in local windows. Across both cutoffs, F statistics fall well below conventional thresholds once the sample is restricted to narrow bandwidths, often below 1 and rarely above 5. By contrast, the global samples deliver strong first stages, with $F > 25$ at the 1,000 cutoff and $F > 130$ at the 2,000 cutoff. This pattern indicates weak identification in local variants and strong relevance only in global specifications. The difference in F statistics across cutoffs also reflects the larger ownership discontinuity at the 2,000 boundary.

Coefficient estimates are correspondingly unstable in local windows. The 2SLS coefficients on index ownership swing widely across narrow bandwidths and often exhibit large standard errors. In several windows the sign flips, which is consistent with weak instruments rather than true economic reversals. By comparison, global estimates are smoother and more interpretable, mirroring their stronger first-stage diagnostics. The instability is especially pronounced for volatility and price informativeness, where local samples are smallest.

Table A.1: IV Robustness Grid: Russell 1000

	50	100	150	200	300	500	global
Amihud illiquidity	0.0744 (0.0563)	0.0067 (0.0220)	0.0924 (0.1595)	0.0050 (0.0370)	0.0474 (0.5740)	0.4971 (2.2070)	0.2319 (0.1450)
	F=2.46 p=0.1169	F=2.16 p=0.1412	F=0.43 p=0.5111	F=2.14 p=0.1432	F=0.02 p=0.8861	F=0.05 p=0.8170	F=27.79 p=1.35e-07
Volatility	2.4288 (3.3057)	3.8671 (3.3886)	13.7755 (22.5574)	4.1938 (4.0521)	-10.8720 (73.8257)	20.9380 (98.3811)	-1.4850*** (0.4090)
	F=2.45 p=0.1172	F=2.19 p=0.1391	F=0.43 p=0.5132	F=2.15 p=0.1423	F=0.02 p=0.8774	F=0.05 p=0.8316	F=27.30 p=1.74e-07
Price informativeness	21.3384 (52.7615)	16.8883 (35.3193)	29.6162 (111.4591)	21.7115 (41.5191)	322.2773 (1071.8428)	109.6254 (543.8555)	-5.8808 (4.0159)
	F=2.42 p=0.1196	F=2.30 p=0.1291	F=0.39 p=0.5339	F=2.26 p=0.1324	F=0.10 p=0.7560	F=0.05 p=0.8225	F=28.49 p=9.44e-08
N	3009	6135	9081	12432	19556	36777	235544
Firms	539	791	950	1068	1276	1761	8627
Dates	99	99	99	99	99	99	99
Clusters	539	791	950	1068	1276	1761	8627
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Firm + time	Firm + time	Firm + time	Firm + time	Firm + time	Firm + time	Firm + time
SE type	Firm clustered	Firm clustered	Firm clustered	Firm clustered	Firm clustered	Firm clustered	Firm clustered

Notes: This table reports the quadratic IV robustness grid for the local-window specifications at the Russell 1000 cutoff. Each cell reports the 2SLS coefficient on *'ind_own'*, its standard error, and the first-stage F statistic / p-value. Bandwidth labels are in basis points; 'global' denotes the unrestricted sample. * p<0.10, ** p<0.05, *** p<0.01.

Weak local-window cases should be interpreted cautiously, especially when the first-stage F statistic is far below the conventional threshold of 10.

Table A.2: IV Robustness Grid: Russell 2000

	50	100	150	200	300	500	global
Amihud illiquidity	-3.9920 (33.3665) F=0.01 p=0.9045	0.6955 (3.8003) F=0.04 p=0.8434	0.0153 (0.2220) F=1.42 p=0.2335	0.4610 (2.0254) F=0.07 p=0.7967	-0.2935 (0.5156) F=0.79 p=0.3739	-0.4358 (0.5759) F=0.81 p=0.3673	-0.2849*** (0.0776) F=131.16 p=<1e-16
Volatility	-0.5604 (22.8979) F=0.01 p=0.9159	1.1773 (13.9026) F=0.05 p=0.8259	-0.7824 (2.2575) F=1.47 p=0.2250	6.9003 (34.4527) F=0.05 p=0.8304	1.3420 (3.8259) F=0.70 p=0.4037	-3.7182 (5.1797) F=0.75 p=0.3859	-0.1767 (0.1263) F=128.98 p=<1e-16
Price informativeness	93.0678 (784.7418) F=0.02 p=0.8933	-108.9001 (520.3633) F=0.05 p=0.8297	-3.9988 (34.0028) F=1.31 p=0.2515	115.0171 (502.9717) F=0.06 p=0.8045	15.6064 (45.6625) F=0.90 p=0.3419	32.3535 (52.5557) F=0.90 p=0.3440	-12.0868*** (1.7568) F=131.12 p=<1e-16
N	4947	10458	15777	21039	30804	49683	235544
Firms	937	1448	1786	2015	2344	2877	8627
Dates	99	99	99	99	99	99	99
Clusters	937	1448	1786	2015	2344	2877	8627
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Firm + time	Firm + time	Firm + time	Firm + time	Firm + time	Firm + time	Firm + time
SE type	Firm clustered	Firm clustered	Firm clustered	Firm clustered	Firm clustered	Firm clustered	Firm clustered

Notes: This table reports the quadratic IV robustness grid for the local-window specifications at the Russell 2000 cutoff. Each cell reports the 2SLS coefficient on 'ind_own', its standard error, and the first-stage F statistic / p-value. Bandwidth labels are in basis points; 'global' denotes the unrestricted sample. * p<0.10, ** p<0.05, *** p<0.01.

Weak local-window cases should be interpreted cautiously, especially when the first-stage F statistic is far below the conventional threshold of 10.

The robustness grid illustrates a standard trade-off between local identification credibility and statistical power. Local windows strengthen the exclusion restriction by comparing nearly identical firms but undermine relevance because the ownership jump becomes small. In this dataset, that relevance loss dominates, making local estimates too noisy for inference. Consequently, the main IV conclusions rely on the global cutoffs, with local results retained as qualitative checks. This approach is consistent with the broader literature on index reconstitution designs that faces similar power limitations.

A.3 QUADRATIC IV R1000/R2000 - FULL RESULTS

This appendix section documents the nonlinear IV robustness checks using the Russell 1000 cutoff. The goal is to evaluate whether the quadratic IV results are stable when the source of exogenous variation is shifted from the 2000 cutoff to the 1000 cutoff. Because the ownership discontinuity at the 1000 threshold is smaller, these specifications are expected to be more fragile. The results are therefore reported as robustness evidence rather than as primary causal estimates.

The nonlinear IV design instruments both ind_own and ind_own^2 using the reconstitution indicator interacted with polynomial functions of the running variable. Controls include firm size and dollar volume as well as the running variable and its square, with firm and time fixed effects. This design mirrors the preferred cutoff-2000 specification but substitutes the 1000 cutoff as the source of identification. The same estimation protocol is applied to facilitate comparability across cutoffs.

The cutoff-1000 nonlinear IV diagnostics are unstable and not economically interpretable. First-stage statistics for ind_own and ind_own^2 become erratic, with signs of numerical instability in parts of the workflow. These problems indicate weak iden-

tification for the quadratic terms at the 1000 cutoff. As a result, coefficient estimates are sensitive to specification choices and should

The instability of the 1000-cutoff quadratic IV design is consistent with a smaller ownership discontinuity at the large-cap boundary. With limited variation in ownership, identifying both the level and square terms becomes especially difficult, amplifying weak-instrument concerns. Consequently, the nonlinear IV evidence relies on the 2000-cutoff specification, which delivers strong first-stage diagnostics. The 1000-cutoff results are retained here as a robustness appendix to demonstrate the sensitivity of nonlinear IV inference to instrument strength.

A.4 MORE TABLES

Table A.3: Quadratic IV Comparison: Russell 1000

	b1	b2	Turning point	In support	F(own)	F(own ²)
Amihud illiquidity	0.4557	-2.7663*	0.0824	Yes	-5878.688	-2867.905
Volatility	-3.0590***	14.6507***	0.1044	Yes	-1994.040	552.483
Price informativeness	0.8048	-71.3398*	0.0056	No	-4238.726	-1945.796

Controls: 'c_firm_size', 'c_dollar_vol', 'running', 'running²'. Firm and time fixed effects. Firm-clustered standard errors. Sample size is not reported in this comparison CSV.

Notes: This table reports the quadratic IV comparison for the global design at the Russell 1000 cutoff. Coefficient entries are reported with significance stars based on p-values. The turning-point indicator reports whether the estimated turning point lies inside the [q01, q99] support band. * p<0.10, ** p<0.05, *** p<0.01. The first-stage columns report the joint F statistics for 'ind_wn' and 'ind_wn²'. *Sample size is not reported in this comparison CSV.*

Table A.4: Quadratic IV Comparison: Russell 2000

	b1	b2	Turning point	In support	F(own)	F(own ²)
Amihud illiquidity	0.3313	-4.8236**	0.0343	Yes	638.786	28.413
Volatility	-2.5336***	15.6492***	0.0809	Yes	611.060	25.592
Price informativeness	-6.8010	-35.9083	-0.0947	No	664.395	34.308

Controls: 'c_firm_size', 'c_dollar_vol', 'running', 'running²'. Firm and time fixed effects. Firm-clustered standard errors. Sample size is not reported in this comparison CSV.

Notes: This table reports the quadratic IV comparison for the global design at the Russell 2000 cutoff. Coefficient entries are reported with significance stars based on p-values. The turning-point indicator reports whether the estimated turning point lies inside the [q01, q99] support band. * p<0.10, ** p<0.05, *** p<0.01. The first-stage columns report the joint F statistics for 'ind_wn' and 'ind_wn²'. *Sample size is not reported in this comparison CSV.*

Table A.5: Quadratic Nonlinearity Results

	Amihud illiquidity	Volatility	Price informativeness
Linear term (b_1)	0.0107 (0.0333)	-0.3250*** (0.0626)	-10.3568*** (1.1805)
Quadratic term (b_2)	0.1882** (0.0947)	0.9991*** (0.2076)	26.7002*** (4.8399)
Turning point	-0.0284 (0.1002)	0.1626 (0.0134)	0.1939 (0.0143)
In support	No	Yes	Yes
Observations	241199	240515	239758
Entities	8877	8858	8855
Time periods	100	100	100
Support min	0.0098	0.0098	0.0098
Support q01	0.0104	0.0104	0.0104
Support q05	0.0128	0.0128	0.0129
Support q50	0.0510	0.0511	0.0513
Support q95	0.1821	0.1822	0.1822
Support q99	0.2267	0.2268	0.2266
Support max	0.9526	0.9526	0.9526

Notes: This table reports the reduced-form quadratic fixed-effects results for the three market-quality outcomes. Coefficients are reported with standard errors underneath in parentheses, and significance stars are based on p-values. The turning-point row reports the implied optimum of the fitted quadratic and its standard error. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The support rows summarize the empirical distribution of ‘*ind_own*’ used to interpret whether the turning point lies inside the observed range.

Table A.6: Spline Panel Regression Results

	Amihud illiquidity	Volatility	Price informativeness
Spline basis 1	-0.0103*** (0.0018)	-0.0022* (0.0013)	-0.1057*** (0.0165)
Spline basis 2	-0.0123*** (0.0017)	-0.0142*** (0.0010)	-0.2848*** (0.0145)
Spline basis 3	-0.0089*** (0.0022)	-0.0266*** (0.0033)	-0.7690*** (0.0323)
Spline basis 4	0.0039 (0.0028)	-0.0129*** (0.0038)	-0.5774*** (0.0409)
Observations	241199	240515	239758
Entities	8877	8858	8855
Time periods	100	100	100
Within R^2	0.0898	0.2200	0.0132
Best ' <i>ind_own</i> '	0.0467	0.0733	0.0104
Best curve value	-0.0002	-0.0050	0.8715
Support q01	0.0104	0.0104	0.0104
Support q50	0.0510	0.0511	0.0513
Support q99	0.2267	0.2268	0.2266

Notes: This table reports spline panel regressions with firm and time fixed effects. Standard errors are reported underneath coefficients in parentheses. The spline basis is a natural cubic spline in '*ind_own*' with four basis terms. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The sample support rows report the empirical ownership distribution used for interpreting the fitted spline.

Table A.7: IV Regression Results for Spline IV at Global Russell 2000 Cutoff

	Amihud illiquidity	Volatility	Price informativeness
Spline basis 1	0.0095 (0.0191)	-0.0086 (0.0269)	-0.3619 (0.3264)
Spline basis 2	0.0095** (0.0038)	-0.0068 (0.0072)	-0.3663*** (0.0964)
Spline basis 3	-0.0040 (0.0440)	0.0343 (0.0781)	-1.0792 (1.0048)
Spline basis 4	-0.0093 (0.0487)	0.0093 (0.0581)	-0.4132 (0.6949)
Observations	235544	234897	234281
Firms	8627	8605	8614
Dates	99	99	99
Cutoff	2000	2000	2000
Poly degree	4	4	4
Min first-stage F	15.7369	15.9077	15.6617
Best ‘ <i>ind_own</i> ’	22.7%	3.3%	14.7%
Controls	Yes	Yes	Yes
Fixed effects	firm + date	firm + date	firm + date
SE type	firm clustered	firm clustered	firm clustered

Notes: This table reports the second-stage IV spline coefficients for the global Russell 2000 design. Standard errors are reported underneath coefficients in parentheses. The dependent-variable panels correspond to Amihud illiquidity, volatility, and price informativeness. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The spline basis is a natural cubic spline in ‘*ind_own*’ with firm and date fixed effects and firm-clustered standard errors.

APPENDIX B

SECOND APPENDIX

B.1 MORE TABLES

Table B.1: Size-Sorted Regression Results by Market Capitalization Decile

(a) Panel A: Amihud illiquidity				(b) Panel B: Volatility			
Decile	Ind. own.	Firm size	Dol. vol.	Decile	Ind. own.	Firm size	Dol. vol.
1	-0.0678 (0.1017)	-0.1330*** (0.0141)	-0.0566*** (0.0027)	1	-0.2098* (0.1089)	-0.1129*** (0.0072)	0.0835*** (0.0037)
2	-0.0264 (0.0193)	-0.0175*** (0.0016)	-0.0173*** (0.0008)	2	-0.0730 (0.0710)	-0.1071*** (0.0072)	0.0769*** (0.0041)
3	0.0191** (0.0084)	-0.0089*** (0.0007)	-0.0079*** (0.0004)	3	-0.0157 (0.0606)	-0.1255*** (0.0077)	0.0805*** (0.0034)
4	0.0007 (0.0027)	-0.0031*** (0.0004)	-0.0042*** (0.0003)	4	-0.1395** (0.0668)	-0.1100*** (0.0058)	0.0790*** (0.0028)
5	-0.0006 (0.0017)	-0.0019*** (0.0002)	-0.0023*** (0.0001)	5	-0.1255** (0.0637)	-0.1234*** (0.0063)	0.0828*** (0.0035)
6	-0.0027 (0.0025)	-0.0006** (0.0003)	-0.0017*** (0.0003)	6	-0.1060** (0.0469)	-0.1037*** (0.0056)	0.0768*** (0.0028)
7	-0.0013 (0.0010)	-0.0002 (0.0002)	-0.0009*** (0.0001)	7	-0.0430 (0.0394)	-0.0960*** (0.0061)	0.0728*** (0.0042)
8	0.0003 (0.0007)	-0.0001 (0.0001)	-0.0005*** (0.0002)	8	0.0407 (0.0544)	-0.0928*** (0.0049)	0.0748*** (0.0034)
9	0.0004 (0.0003)	-0.0000 (0.0000)	-0.0003*** (0.0001)	9	-0.0449 (0.0520)	-0.0811*** (0.0044)	0.0769*** (0.0035)
10	0.0002 (0.0001)	0.0000 (0.0000)	-0.0001*** (0.0000)	10	0.0414 (0.0374)	-0.0620*** (0.0044)	0.0699*** (0.0048)
Observations within R^2		23929 0.2542		Observations within R^2		23929 0.2674	

(c) Panel C: Price informativeness

Decile	Ind. own.	Firm size	Dol. vol.
1	-13.5969*** (1.2993)	-0.5095*** (0.0474)	-0.0180 (0.0149)
2	-8.2094*** (0.6583)	-0.6022*** (0.0537)	0.2268*** (0.0177)
3	-3.8004*** (0.5556)	-0.6176*** (0.0713)	0.2643*** (0.0219)
4	-3.5096*** (0.4750)	-0.5963*** (0.0642)	0.3350*** (0.0240)
5	-3.7967*** (0.5085)	-0.5927*** (0.0607)	0.4377*** (0.0296)
6	-2.4811*** (0.4838)	-0.4569*** (0.0637)	0.4591*** (0.0283)
7	-2.4157*** (0.4029)	-0.2969*** (0.0580)	0.4501*** (0.0288)
8	-2.8178*** (0.5606)	-0.3852*** (0.0548)	0.4437*** (0.0348)
9	-2.2610*** (0.6090)	-0.3975*** (0.0538)	0.4666*** (0.0451)
10	-0.4039 (0.8068)	-0.4402*** (0.0620)	0.4068*** (0.0546)
Observations within R^2		23929 0.0714	

Notes: This table reports the size-sorted regression results in three panels. Standard errors are reported underneath coefficients in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Sample construction: deciles are based on market capitalization. Regressions are estimated separately within each decile for each outcome variable.

Table B.3: Heterogeneity by Recession: Panel FE with Interaction

	Amihud illiquidity	Volatility	Price informativeness
Industry ownership	0.0800*** (0.0237)	-0.0681*** (0.0231)	-3.2530*** (0.2401)
Industry ownership \times recession	-0.2142** (0.0867)	0.1390 (0.1036)	-0.5590 (0.8015)
Firm size	0.0082*** (0.0016)	-0.0983*** (0.0031)	-0.3411*** (0.0267)
Dollar volume	-0.0221*** (0.0017)	0.0747*** (0.0024)	0.1456*** (0.0147)
Observations	241199	240515	239758
Entities	8877	8858	8855
Time periods	100	100	100
Firm FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Two-way clustered SE	Yes	Yes	Yes

Notes: This table reports panel fixed-effects regressions with an interaction between industry ownership and recession periods. Standard errors are reported underneath coefficients in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The interaction coefficient is the incremental slope in recession quarters relative to non-recession quarters. Standard errors are two-way clustered by firm and date.

Table B.4: Business-Cycle by Size-Decile Regression Results: Amihud illiquidity

Size decile	Exp. 1993-2000	Rec. 2001	Exp. 2002-2007	Rec. 2007-2009	Exp. 2009-2019	Rec. 2019-2020	Exp. 2020-2023
1	—	—	4.1130 (10.2765)	—	-6.7137 (9.1838)	-5.4107 (14.0474)	0.4476 (1.3706)
2	0.6909 (2.0553)	-60.5637 (55.0788)	-6.0882*** (1.6093)	-35.2844 (28.3178)	0.0602 (0.9387)	-1.1240 (1.6690)	-0.1624 (0.4595)
3	-1.8919 (5.8612)	-18.7479* (9.7675)	-0.6591 (0.8669)	1.8330 (4.0609)	-0.3110 (0.2847)	-1.7923 (1.2064)	0.0948 (0.1564)
4	-0.5965 (0.9196)	-2.2226 (3.1872)	0.3133 (0.2099)	-0.0062 (0.4088)	-0.1050** (0.0417)	-0.2233 (0.2567)	-0.2077** (0.0980)
5	0.8900 (1.8452)	0.2236 (0.9405)	0.0904 (0.0714)	-0.1004 (0.1287)	-0.0560* (0.0300)	-0.0944 (0.1477)	-0.0315 (0.0361)
6	0.0851 (0.4276)	0.4979 (0.5302)	0.0947** (0.0414)	-0.0479 (0.0531)	0.0016 (0.0111)	-0.0641 (0.0558)	-0.0225 (0.0140)
7	-0.0483 (0.1061)	-0.0231 (0.0388)	0.0348*** (0.0097)	-0.0230 (0.0168)	-0.0008 (0.0022)	-0.0044 (0.0091)	-0.0115*** (0.0041)
8	-0.0857** (0.0356)	0.0159 (0.0263)	0.0095** (0.0045)	-0.0076 (0.0053)	-0.0015 (0.0035)	-0.0122* (0.0067)	-0.0077*** (0.0028)
9	0.0318 (0.0287)	0.0381 (0.0303)	-0.0003 (0.0023)	0.0031 (0.0028)	0.0002 (0.0006)	-0.0028 (0.0032)	-0.0031 (0.0024)
10	0.0079* (0.0047)	0.0013 (0.0021)	-0.0015*** (0.0005)	-0.0002 (0.0008)	0.0002 (0.0001)	-0.0004 (0.0012)	0.0004 (0.0003)

Notes: Each cell reports the coefficient on industry ownership from a two-way fixed-effects regression estimated within a business-cycle segment and size decile. Standard errors are reported underneath coefficients in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The cycle labels are abbreviated as expansion (Exp.) and recession (Rec.). Cell-level sample sizes, entity counts, and time-period counts vary across regressions and are stored in the source CSV. A dash indicates that the source regression cell is not available.

Table B.5: Business-Cycle by Size-Decile Regression Results: Volatility

Size decile	Exp. 1993-2000	Rec. 2001	Exp. 2002-2007	Rec. 2007-2009	Exp. 2009-2019	Rec. 2019-2020	Exp. 2020-2023
1	—	—	-1.6504 (4.7741)	—	1.3217 (1.9270)	-1.3644 (2.5443)	0.0181 (0.4618)
2	—	-4.4391 (12.4485)	-1.1188 (0.8961)	-13.3418** (5.7042)	-0.1360 (0.3636)	-0.2176 (1.5972)	0.5831 (0.4564)
3	1.3540 (6.5851)	-9.5938* (4.8497)	0.3931 (0.6450)	0.4814 (4.2206)	-0.1290 (0.1583)	-1.6612* (0.9987)	0.2015 (0.2941)
4	-0.4578 (1.1001)	3.1674 (4.6317)	-0.0971 (0.3556)	1.8182 (1.3683)	-0.0312 (0.1096)	-0.3983 (0.9718)	-0.0193 (0.2513)
5	0.6315 (3.8147)	-0.4871 (3.2858)	-0.1974 (0.2272)	-0.0011 (0.6172)	-0.0528 (0.1171)	0.1296 (1.0930)	-0.0433 (0.1787)
6	0.3384 (2.5666)	1.4813 (4.1616)	-0.0439 (0.1501)	0.1666 (0.6170)	-0.1032 (0.0951)	-1.7183 (1.2327)	-0.3089** (0.1408)
7	-0.9261 (1.2022)	-1.8811*** (0.4220)	0.0802 (0.1470)	0.2268 (0.4083)	-0.1376** (0.0571)	0.3013 (0.6618)	-0.5120* (0.2910)
8	-0.4373 (1.9467)	-1.2190 (0.8966)	0.1818 (0.1580)	-0.0100 (0.3470)	-0.0643 (0.0520)	-0.7590** (0.3415)	-0.2931* (0.1532)
9	1.6041 (2.3656)	-0.9216 (3.1151)	0.4801** (0.1905)	0.1322 (0.4928)	0.0685* (0.0371)	0.4798 (0.4240)	-0.2146*** (0.0737)
10	3.2393** (1.5136)	-0.5515 (1.7810)	0.2146 (0.1995)	0.9055 (0.8097)	0.0312 (0.0304)	-0.0823 (0.2779)	0.0026 (0.0597)

Notes: Each cell reports the coefficient on industry ownership from a two-way fixed-effects regression estimated within a business-cycle segment and size decile. Standard errors are reported underneath coefficients in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The cycle labels are abbreviated as expansion (Exp.) and recession (Rec.). Cell-level sample sizes, entity counts, and time-period counts vary across regressions and are stored in the source CSV. A dash indicates that the source regression cell is not available.

Table B.6: Business-Cycle by Size-Decile Regression Results: Price informativeness

Size decile	Exp. 1993-2000	Rec. 2001	Exp. 2002-2007	Rec. 2007-2009	Exp. 2009-2019	Rec. 2019-2020	Exp. 2020-2023
1	—	—	-61.4089 (51.0435)	—	6.0464 (27.7230)	50.3437* (25.3792)	-11.0668 (7.5884)
2	-19.0526*** (1.7280)	-276.8213*** (58.8438)	-7.4068 (8.3050)	6.0342 (45.9908)	-12.9812 (9.6485)	24.0119 (20.8303)	1.8399 (5.3546)
3	9.4201 (30.5327)	70.0981 (68.2479)	-3.6210 (18.0952)	1.3959 (34.1508)	-18.5138*** (2.9544)	11.4248 (11.0397)	-8.3910* (4.5077)
4	13.2106 (12.5583)	-22.9894 (84.6787)	-27.4725*** (7.7381)	-22.0223 (23.5537)	-18.0537*** (1.9969)	-2.8037 (4.4130)	-11.7401*** (3.1138)
5	33.4625 (34.2266)	17.8546 (32.5824)	-12.7003*** (4.8500)	-8.7552 (8.1478)	-8.9452*** (1.1354)	-5.1195 (6.4164)	-11.0855*** (1.4486)
6	37.3708 (30.5184)	-25.3137 (33.4932)	-5.5851*** (2.1343)	-4.0640 (3.9576)	-4.1061*** (0.8037)	6.0145 (5.9116)	-5.3036*** (1.5333)
7	37.2193 (28.9148)	2.8618 (11.2145)	-3.6360 (2.2811)	-1.0371 (3.3058)	-4.0113*** (0.7319)	-0.7917 (5.9930)	-3.3728** (1.5294)
8	9.4079 (26.5372)	-1.2006 (10.5442)	-14.5853*** (2.7091)	-2.0700 (3.2440)	-3.0247*** (0.6342)	1.1919 (2.9233)	-1.2936 (0.9958)
9	-12.1230 (17.4200)	12.0239 (48.6792)	-16.3989*** (2.9173)	0.5630 (3.0757)	-2.0913*** (0.6218)	-2.0598 (2.6029)	-2.1893*** (0.7196)
10	-12.9384 (22.8037)	-34.1166 (39.1784)	-14.0360*** (3.7638)	-5.0692 (5.5135)	-1.7743** (0.7481)	2.8694 (3.3778)	-2.1335** (0.9855)

Notes: Each cell reports the coefficient on industry ownership from a two-way fixed-effects regression estimated within a business-cycle segment and size decile. Standard errors are reported underneath coefficients in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

The cycle labels are abbreviated as expansion (Exp.) and recession (Rec.). Cell-level sample sizes, entity counts, and time-period counts vary across regressions and are stored in the source CSV. A dash indicates that the source regression cell is not available.

Table B.7: Best Business-Cycle by Size-Decile Results

Panel A: Global top 15 significant cells

Rank	Dep. var.	Cycle	Decile	β (SE)	p	Score
1	Price informativeness	Exp. 2009-2019	4	-18.0537***	0.0000	80.00
2	Price informativeness	Exp. 2009-2019	5	-8.9452***	0.0000	72.07
3	Price informativeness	Exp. 2020-2023	5	-11.0855***	0.0000	61.22
4	Price informativeness	Exp. 1993-2000	2	(1.7280)	0.0000	55.54
5	Price informativeness	Exp. 2009-2019	7	(0.7319)	0.0000	52.29
6	Price informativeness	Exp. 2009-2019	3	(2.9544)	0.0000	50.48
7	Price informativeness	Exp. 2002-2007	9	(2.9173)	0.0000	49.91
8	Price informativeness	Exp. 2002-2007	8	(2.7091)	0.0000	48.49
9	Price informativeness	Exp. 2009-2019	6	(0.8037)	0.0000	47.63
10	Price informativeness	Exp. 2009-2019	8	(0.6342)	0.0000	46.60
11	Price informativeness	Exp. 2009-2019	9	(0.6218)	0.0008	33.53
12	Price informativeness	Exp. 2002-2007	10	(3.7638)	0.0002	33.33
13	Amihud illiquidity	Exp. 2002-2007	7	(0.0097)	0.0003	32.37
14	Volatility	Rec. 2001	7	-1.8811***	0.0000	30.33
15	Price informativeness	Exp. 2020-2023	4	(3.1138)	0.0002	29.42

Panel B: Top 5 significant cells by dependent variable

Rank	Dep. var.	Cycle	Decile	β (SE)	p	Score
<i>Amihud illiquidity</i>						
1	Exp.	2002-2007	7	0.0348***	0.0003	32.37
2	Exp.	2002-2007	10	-0.0015***	0.0045	25.39
3	Exp.	2020-2023	8	-0.0077***	0.0067	24.10
4	Exp.	2020-2023	7	(0.0041)	0.0052	23.94
5	Exp.	2002-2007	2	(1.6093)	0.0002	23.93
<i>Volatility</i>						
6	Rec.	2001	7	-1.8811***	0.0000	30.33
7	Exp.	2020-2023	9	-0.2146***	0.0036	26.84
8	Exp.	2009-2019	7	(0.0571)	0.0159	23.00
9	Exp.	2002-2007	9	(0.1905)	0.0118	22.38
10	Exp.	2020-2023	6	(0.1408)	0.0283	18.23
<i>Price informativeness</i>						
11	Exp.	2009-2019	4	(1.9969)	0.0000	80.00
12	Exp.	2009-2019	5	(1.1354)	0.0000	72.07
13	Exp.	2020-2023	5	(1.4486)	0.0000	61.22
14	Exp.	1993-2000	2	(1.7280)	0.0000	55.54
15	Exp.	2009-2019	7	(0.7319)	0.0000	52.29

Notes: The table lists the highest-scoring significant coefficient estimates on industry ownership. Score is defined in the source workflow as $|t| \times \log(1 + N)$. Standard errors are reported underneath coefficients in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The global panel ranks all significant cells, while the dependent-variable panel keeps the top five cells within each outcome group.

BIBLIOGRAPHY

- [1] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.
- [2] E.T. Jaynes. Probability theory as logic. In Fougère, P.F., editor, *Maximum entropy and Bayesian methods*, pages 1–16. Kluwer, Dordrecht, 1990. URL <http://bayes.wustl.edu/etj/articles/prob.as.logic.pdf>. Proceedings of the 9th International Workshop, Dartmouth.