POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

EXECUTIVE SUMMARY OF THE THESIS

The Economic Consequences of Shareholder Breadth: Evidence on Operating Performance and Market Valuation

TESI MAGISTRALE IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

AUTHOR: Tommaso Rossi

ADVISOR: Benedetta Montanaro

CO-ADVISOR: Andriy Bodnaruk

ACADEMIC YEAR: 2024 - 2025

1. Introduction

Capital markets are characterized by persistent informational frictions. Even in developed financial systems, investors do not observe or actively follow all available securities, as information acquisition is costly and attention is limited. Consequently, firms differ substantially in the breadth of their shareholder base: while some companies are widely held and recognized by a large and diversified investor population, others remain relatively neglected and characterized by concentrated ownership structures. Such heterogeneity in investor recognition has important implications for asset pricing and corporate finance.

The theoretical foundation of this thesis is grounded in Merton's (1987) model of capital market equilibrium with incomplete information. In this framework, investors are aware of only a subset of securities and construct portfolios based on limited recognition sets. Because idiosyncratic risk can only be shared among investors who recognize a given security, firms followed by a broader investor base benefit from improved risk sharing. This mechanism reduces the equilibrium required rate of return and increases firm valuation. Conversely, firms with limited recognition must compensate investors with higher expected returns, resulting in a valuation discount. The model therefore predicts that investor recognition is a priced determinant of firm value and that differences in shareholder breadth may generate persistent variation in valuation across firms.

Subsequent empirical literature on neglected firms, ownership breadth and investor attention has provided evidence broadly consistent with these theoretical predictions. Firms with limited visibility often exhibit lower liquidity, reduced analyst coverage and higher expected returns. However, empirical findings remain heterogeneous, particularly when extending the analysis beyond cross-sectional pricing effects to

examine dynamic relationships within firms over time. In addition, relatively less attention has been devoted to the broader implications of shareholder breadth for real economic outcomes.

Against this background, this thesis addresses two central research questions:

- Is a broader shareholder base associated with higher firm valuation, consistent with incomplete information theory?
- Is a broader shareholder base associated with improved operating performance?

The empirical analysis is conducted on a firm-year panel built from Compustat Daily Updates – Annual Fundamentals for North American firms over the period 2000–2024. Investor recognition is proxied by the number of common shareholders, while firm outcomes are measured through market-based valuation indicators and accounting-based performance measures. The empirical framework relies on panel regressions with firm and year fixed effects, allowing identification to be based on within-firm variation over time and mitigating concerns related to time-invariant unobserved heterogeneity.

The dataset is structured to support a symmetric investigation of investor recognition across two complementary outcomes: operating performance and market valuation. Operating performance is measured by return on assets (ROA), while market valuation is measured by Tobin's Q, allowing the analysis to distinguish between realized profitability and the market's expectations embedded in prices.

The non-parametric results reveal a positive pooled association between shareholder breadth and operating profitability and a negative pooled association with Tobin's Q. In fixed-effects models, the profitability effect weakens substantially once time-varying fundamentals are controlled for, while the valuation results display a more nuanced pattern: breadth is negatively associated with Q in baseline specifications but becomes positively and convexly related to valuation once firm scale and capital structure are accounted for. Overall, the findings indicate that investor recognition plays a role for market valuation after conditioning on firm fundamentals, whereas its link to operating performance remains limited.

2. Variable construction and sample preparation

Investor recognition is proxied by shareholder breadth, measured using the Compustat shareholder-count series (cshr). The panel is constructed to preserve meaningful within-firm variation over time while limiting distortions from missing observations. Shareholder counts are converted into absolute values and, due to strong right-skewness, enter the parametric specifications predominantly in logarithmic form to improve stability in fixed-effects estimation.

Firm outcomes are examined along two complementary dimensions. Operating performance is measured by return on assets (ROA), computed as net income scaled by total assets and winsorized to mitigate the influence of extreme observations. Market valuation is captured by Tobin's Q, constructed as a market-to-book style ratio and transformed into logarithmic form to address skewness. This symmetric treatment, winsorization for profitability and logarithmic transformation for valuation, ensures statistical stability across outcome measures.

Time-varying firm fundamentals are included as controls in the fixed-effects regressions. Firm size is measured as the logarithm of total assets and leverage as total debt scaled by total assets. Logarithmic transformations for breadth, Tobin's Q, and size, combined with winsorization of ROA, improve numerical conditioning and prevent results from being driven by scale effects or extreme realizations.
Additional variables, including cash holdings and related balance-sheet aggregates, were also constructed and tested in preliminary specifications; however, they were not retained in the final regressions as they did not provide incremental explanatory power once the core controls were included.

3. Non-Parametric Evidence

Before proceeding to parametric fixed-effects estimation, the analysis develops a non-parametric assessment of the association between shareholder

breadth and firm outcomes. This preliminary step provides a distribution-free evaluation of monotone relationships that is robust to heavy tails, nonlinear rescaling and extreme observations. The procedure relies on Spearman's rank correlation coefficient combined with permutation-based inference (B = 2000), ensuring that statistical significance does not depend on parametric assumptions.

For completeness, the analysis begins with the unconditional test pairing shareholder count in levels (cshr_abs) with raw ROA. Using all firm-years with finite values (n = 123,289), the observed Spearman correlation equals ϱ_S = 0.1845.
The test is then replicated using the treated variables employed in the parametric models, pairing the logarithmic breadth measure (breadth_no1) with the winsorized profitability variable (roa_w). Because Spearman's coefficient depends only on ranks and both transformations are monotone on the interior support, the result remains almost unchanged, with ϱ_S = 0.2195. The persistence of the magnitude confirms that the positive monotonic association between recognition and operating profitability is not driven by extreme values or scaling choices.

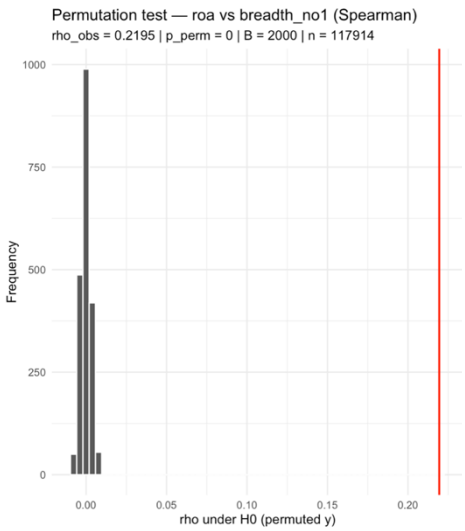

Figure 1 - Permutation test – ROA winsorized vs breadth_no1

An analogous procedure is implemented for valuation. The initial test pairs Tobin's Q in levels with shareholder count in levels (cshr_abs), yielding ϱ_S = −0.1347 based on n = 112,201 firm-years. The negative sign indicates that, in pooled data, firms with broader shareholder bases tend to rank lower in the cross-sectional valuation distribution.
The analysis is then repeated using the treated variables employed in the parametric specifications, namely the logarithmic transformation of Tobin's Q (q_log_w) and the log-transformed breadth proxy (breadth_no1). The result remains highly consistent, with ϱ_S = −0.1370 on n = 111,997 observations. The similarity in magnitude across raw and treated versions confirms that the negative unconditional association is not an artifact of heavy tails or skewness adjustments.

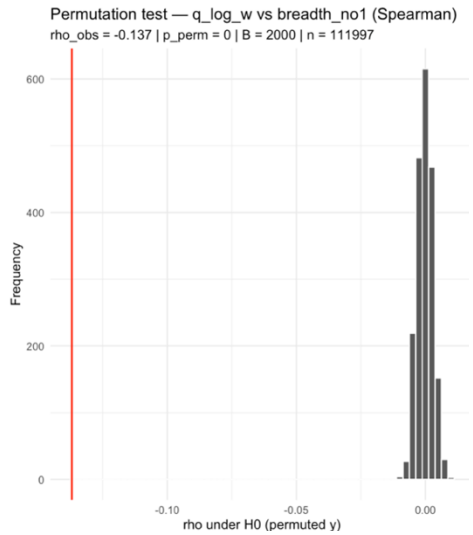

Figure 2 - Permutation test – Log1p of Tobin's Q winsorized vs breadth_no1

4. Parametric Evidence: Two-Way Fixed Effects (Breadth-Only Specifications)

After establishing unconditional patterns through rank-based diagnostics, the analysis proceeds with parametric estimation using a two-way fixed-effects framework estimated using the feols function. The objective is to identify within-firm associations over time while removing time-invariant heterogeneity and common macroeconomic shocks.

The baseline econometric specification is:

$$Y\_it = \alpha\_i + \tau\_t + \beta 1 * breadth\_no1\_it + u\_it$$

where $Y\_\{it\}$ denotes the outcome of interest (either ROA_w or q_log_w) for firm i in year t; α_i are firm fixed effects; τ_it are year fixed effects;

and $u\_\{it\}$ is the error term clustered at the firm level.

Firm fixed effects absorb all time-invariant characteristics of each company, including business model, persistent industry positioning, managerial quality and structural profitability differences. Year fixed effects absorb aggregate shocks common to all firms, such as macroeconomic fluctuations, regulatory changes or market-wide valuation cycles. Identification therefore relies exclusively on within-firm variation: the coefficient $\beta 1$ captures whether years in which a given firm experiences higher shareholder breadth are also years in which it exhibits higher profitability or valuation relative to its own long-run average.
To allow for potential curvature while maintaining parsimony, a centered quadratic specification is also estimated:

$$Y\_it = \alpha\_i + \tau\_t + \gamma 1 * (breadth\_no1\_it - \mu\_b) + \gamma 2 * (breadth\_no1\_it - \mu\_b)\text{^}2 + u\_it$$

where μ_b denotes the sample mean of breadth on the estimation sample. Centering ensures that $\gamma 1$ measures the marginal effect at the center of support and reduces collinearity between linear and quadratic terms.
In addition, flexible spline specifications are estimated to relax global polynomial restrictions. Splines approximate the relationship through piecewise polynomial segments connected at predefined knots, allowing the marginal effect of shareholder breadth to vary locally across different regions of the distribution without imposing symmetry or a single turning point. This approach is particularly relevant in the context of incomplete-information models, which do not necessarily imply constant marginal recognition effects across the participation spectrum. Recognition gains may be stronger at intermediate levels, where additional investors materially improve risk sharing or may flatten once ownership becomes sufficiently diffuse.
Across all specifications (linear, quadratic and spline) the identifying variation remains strictly within-firm over time. The comparison of functional forms allows the analysis to assess whether the recognition effect is approximately linear, exhibits convexity or concavity or concentrates in specific segments of the breadth distribution.

4.1. Fixed Effects: ROA

The fixed-effects analysis for operating performance uses the pooled winsorized profitability measure (roa_w) as the dependent variable and the logarithmic breadth proxy (breadth_no1) as the key regressor. The baseline linear specification is estimated on 117,914 firm-years after excluding single-shareholder observations and requiring finite outcomes.
The linear model yields a positive and statistically highly significant coefficient:

	Linear FE		
	Est.	Std. Error	p-value
breadth_no1	0.040***	0.008	<0.001
Num.Obs.	117914		
RMSE	1.64		
R2	0.559		
R2 Adj.	0.510		

Table 1 - Results of linear fixed-effects regression (ROA, breadth only)

The magnitude implies that, within the same firm and after absorbing year effects, a one-unit increase in log shareholder breadth is associated with an increase of approximately 0.04 in winsorized ROA. Although the statistical significance is strong, the explanatory power remains limited, indicating that recognition explains only a small fraction of within-firm profitability variation (within R^2 7.327e-4).
To allow for global curvature, a centered quadratic specification is estimated. The linear term is still positive and significant ($\gamma_1 = 0.042$, $p < 0.001$), while the quadratic term is negative and non-significant ($\gamma_2 = -0.003$, $p < 0.055$). The implied turning point occurs at breadth_no1 ≈ 9.9.13, located near the extreme upper tail of the empirical distribution.
The spline terms jointly provide a similarly small but non-zero increase in within R^2, the second basis coefficient is positive and significant at the five percent level and the first and third basis terms are not statistically different from zero at conventional thresholds. RMSE is again 1.641. This pattern aligns with the centered quadratic, namely a positive slope over the interior of the distribution with mild flattening in the far right tail.

4.2. Fixed Effects: Tobin's Q

The valuation analysis mirrors the profitability framework but uses the logarithmic transformation of Tobin's Q (q_log_w) as the dependent variable. As in the ROA case, estimation is conducted within a two-way fixed-effects framework, absorbing firm and year fixed effects and clustering standard errors at the firm level. The identifying variation therefore comes from within-firm changes in shareholder breadth over time.

In the baseline linear specification, the coefficient on breadth_no1 is negative and statistically significant ($\beta = -0.015$, SE = 0.003, $p < 0.001$). The magnitude indicates that, within firms, years characterized by higher shareholder breadth are associated with slightly lower valuation multiples relative to the firm's historical average. Although economically modest, the effect is precisely estimated. The within R^2 equals 0.711.

	Linear FE		
	Est.	Std. Error	p-value
breadth_no1	−0.015***	0.003	<0.001
Num.Obs.	111997		
RMSE	0.45		
R2	0.711		
R2 Adj.	0.677		

Table 2 - Results of linear fixed-effects regression (log Tobin's Q, breadth only)

The quadratic specification confirms the presence of mild curvature. The linear centered term is negative and significant ($\beta_1 = -0.018$, $p < 0.001$), while the squared term is positive and significant ($\beta_2 = 0.003$, $p < 0.001$). This implies a convex relationship: at lower levels of recognition, increases in breadth are associated with declines in Q, but the marginal effect attenuates and eventually becomes positive at sufficiently high levels of breadth. The turning point lies toward the upper region of the empirical distribution.
Spline specifications (df = 3) broadly support this pattern. The first spline segment is negative and statistically significant (−0.336, $p = 0.032$), while subsequent segments are statistically weaker. Overall goodness-of-fit remains unchanged ($R^2 = 0.711$), indicating that allowing flexible curvature does not materially improve explanatory power relative to the quadratic model.

5. Parametric Evidence: Two-Way Fixed Effects with Controls

The second step of the parametric analysis augments the baseline two-way fixed-effects framework with time-varying balance-sheet controls. The purpose of this extension is to address the possibility that changes in shareholder breadth co-move with changes in firm scale and capital structure, which are themselves strongly related to both operating performance and market valuation.
The controlled linear specification is defined as:

$$Y_{it} = \alpha_i + \tau_t + \beta 1 * breadth\_c + \beta 2 * size\_c\_it + \beta 3 * liab\_c\_it + u\_it$$

where breadth_c, size_c and liab_c denote the centered logarithmic breadth, total assets and total liabilities measures. Centering is performed on the estimation sample means to improve numerical conditioning and facilitate interpretation of marginal effects at the center of support.
As in the breadth-only specifications, firm and year fixed effects are included and standard errors are clustered at the firm level. Identification therefore continues to rely exclusively on within-firm variation over time. The inclusion of scale and leverage controls allows the analysis to distinguish between a direct recognition channel and indirect effects operating through contemporaneous changes in firm size and capital structure.
To allow for nonlinearities in the presence of controls, quadratic extensions are also estimated for breadth and size, and spline specifications are introduced as flexible robustness checks. Across all controlled models, the fixed-effects structure remains unchanged.

5.1. Fixed Effects: ROA (with Controls)

Relative to the breadth-only models, the inclusion of size and leverage dramatically increases explanatory power: the within R^2 rises to 0.351 and the RMSE declines to 1.319, indicating that time-varying balance-sheet conditions account for a substantial share of within-firm profitability movements. The control variables enter with economically intuitive and highly significant coefficients. Firm size exhibits a strong positive

association with operating profitability while leverage is strongly negative.
Crucially, once these controls are included, the coefficient on shareholder breadth sign is negative and statistically significant.

	Linear FE + controls		
	Est.	Std. Error	p-value
breadth_c	−0.033***	0.007	<0.001
size_c	1.542***	0.031	<0.001
liab_c	−0.916***	0.024	<0.001
Num.Obs.	117536		
RMSE	1.32		
R2	0.715		
R2 Adj.	0.683		

Table 3 - Results of linear fixed-effects regression with controls (ROA)

This sign reversal relative to the linear breadth-only specifications indicates that the previously documented positive association between recognition and profitability was largely driven by concurrent changes in firm scale and leverage. Conditional on these balance-sheet dynamics, the partial within-firm association between shareholder breadth and ROA is small and adverse on average.
To assess curvature in the controlled setting, quadratic terms are introduced for breadth and size. The within R^2 increases further to 0.438 and RMSE declines to 1.228, confirming the dominant explanatory role of scale effects. The size coefficients display clear concavity ($\delta_1 = 0.8117$, $\delta_2 = -0.08663$, both highly significant), while leverage remains strongly negative ($\beta_3 = -0.6067$).In contrast, the breadth linear and quadratic terms become statistically indistinguishable from zero ($\gamma_1 = -0.00546$, $p = 0.376$; $\gamma_2 = -0.000519$, $p = 0.669$), indicating that once time-varying size and leverage are absorbed, a flexible quadratic in breadth does not contribute explanatory power for operating performance.
Spline specifications yield qualitatively identical conclusions: allowing flexible local curvature in breadth does not materially alter the fit or the inference and the recognition effect remains economically negligible once core balance-sheet controls are included.

5.2. Fixed Effects: Tobin's Q (with Controls)

The controlled valuation specification augments the baseline two-way fixed-effects framework with time-varying size and leverage controls, allowing the analysis to distinguish the direct recognition channel from contemporaneous balance-sheet dynamics. Estimation is conducted on the complete-regressor sample and continues to rely exclusively on within-firm variation over time.

In the linear controlled model, the coefficient on centered breadth becomes positive and statistically significant ($\beta = 0.010$, SE = 0.003, $p = 0.001$). The inclusion of controls increases explanatory power materially ($R^2 = 0.763$; RMSE declines from 0.45 to 0.41). This sign reversal relative to the breadth-only specification indicates that the negative association previously observed was largely driven by omitted scale and capital-structure effects.
Interestingly, size enters negatively ($\beta = -0.262$, $p < 0.001$), suggesting that within firms, expansions in asset scale are associated with lower valuation multiples once fixed effects are absorbed. Leverage enters positively but modestly ($\beta = 0.013$, $p = 0.037$), indicating a small conditional association with valuation in this linear specification.
The quadratic controlled model reveals a more nuanced structure. The linear breadth term becomes statistically insignificant ($\beta_1 = -0.001$, $p = 0.628$), while the squared term is positive and significant ($\beta_2 = 0.002$, $p = 0.003$). This confirms a convex recognition–valuation relationship once scale effects are held constant. The marginal effect of breadth is weak around the sample mean but increases at higher levels of recognition. Model fit improves further ($R^2 = 0.786$; RMSE = 0.39).

	Quadratic FE + controls		
	Est.	Std. Error	p-value
breadth_c	−0.001	0.003	0.628
I(breadth_c^2)	0.002**	0.001	0.003
size_c	−0.059***	0.007	<0.001
I(size_c^2)	0.024***	0.001	<0.001
lev_c	−0.075***	0.006	<0.001
Num.Obs.	111660		
RMSE	0.39		
R2	0.786		
R2 Adj.	0.761		

Table 4 - Results of quadratic fixed-effects regression with controls (log Tobin's Q)

Spline specifications with controls corroborate this convex pattern. The first spline segment is negative and significant (−0.281, p = 0.029), while subsequent segments turn positive, although with lower precision. Goodness-of-fit remains stable relative to the quadratic model (R^2 = 0.786), indicating that the flexible functional form does not fundamentally alter inference.

6. Conclusion

This thesis investigates whether investor recognition, proxied by shareholder breadth, is economically relevant for firm valuation and operating performance.

The non-parametric evidence reveals a positive monotonic association between shareholder breadth and operating profitability and a negative association between breadth and Tobin's Q in pooled data. These unconditional patterns primarily reflect cross-sectional sorting: mature, large-cap firms tend to exhibit broader shareholder bases but lower growth opportunities, while younger, high-growth firms often display narrower ownership structures and higher valuation multiples.

When firm and year fixed effects are introduced, the recognition–valuation relationship evolves in a more nuanced manner. In breadth-only fixed-effects specifications, increases in shareholder breadth remain negatively associated with Tobin's Q within firms, although the relationship displays convexity, with the marginal effect attenuating at higher levels of recognition. This pattern suggests that lifecycle and maturity dynamics continue to influence valuation even after removing time-invariant heterogeneity. However, once time-varying size and leverage controls are included, the sign of the linear breadth coefficient reverses and becomes positive. The controlled quadratic and spline models further reveal a convex relationship, with the marginal valuation impact strengthening at higher levels of breadth. The consistency of this pattern across functional forms indicates that the result is not driven by specification choices. Economically, these findings suggest that the negative association observed in simpler models largely reflects omitted time-varying fundamentals, while the positive convex effect emerging after conditioning is consistent with incomplete-information theory: broader investor recognition enhances risk sharing and supports higher valuation multiples once structural firm characteristics are properly accounted for.

The evidence for operating performance is more nuanced. In breadth-only fixed-effects models, increases in shareholder breadth are associated with modest improvements in ROA. However, once scale and leverage dynamics are controlled for, the direct association between breadth and profitability weakens substantially and becomes economically negligible. This suggests that the positive unconditional relationship between recognition and profitability largely reflects concurrent changes in firm size and capital structure rather than a direct operational channel.

Taken together, the results indicate that investor recognition plays a meaningful role in shaping firm valuation once time-varying fundamentals are explicitly controlled for, while its link to operating performance remains limited after accounting for structural firm characteristics. The findings therefore provide qualified support for the core prediction of incomplete-information models in the domain of market valuation, while highlighting that improvements in shareholder breadth do not automatically translate into higher contemporaneous profitability.

From an academic perspective, this thesis contributes to the literature on investor recognition by clarifying the empirical conditions under which the recognition effect emerges. The analysis shows that unconditional correlations and baseline fixed-effects specifications may deliver misleading inferences about the recognition–valuation relationship. By distinguishing cross-sectional composition effects from fully conditioned within-firm dynamics, the study reconciles seemingly contradictory patterns in the data and demonstrates that the positive valuation effect predicted by incomplete-information models becomes visible only after controlling for time-varying firm fundamentals. Moreover, the evidence of convexity in the recognition–valuation link refines the theoretical implications of Merton's framework, suggesting that the benefits of broader investor participation strengthen once a sufficiently wide shareholder base is achieved.

Methodologically, the thesis highlights the importance of specification choice in panel settings. The results illustrate how pooled correlations, baseline fixed-effects regressions and fully controlled models can yield different signs and economic interpretations. By systematically comparing linear, quadratic and spline specifications within a two-way fixed-effects framework, the analysis demonstrates that functional-form assumptions materially affect inference. The documented sign reversal in Tobin's Q once size and leverage are included underscores the necessity of accounting for time-varying fundamentals when studying recognition effects. More broadly, the study provides a transparent empirical framework that can be applied to future research on investor attention, ownership structure, and valuation dynamics.

Beyond its theoretical and methodological relevance, the findings carry practical implications for corporate decision-makers and market participants. For firms and investor-relations functions, the results suggest that expanding the shareholder base may enhance market valuation, particularly once a sufficiently broad recognition level is reached. However, the absence of a robust effect on operating profitability indicates that visibility and investor outreach strategies are unlikely to substitute for improvements in underlying operational performance. For investors, the analysis emphasizes the importance of distinguishing lifecycle and scale effects from genuine recognition-driven valuation adjustments.

References

[1] R. C. Merton, "A Simple Model of Capital Market Equilibrium with Incomplete Information," The Journal of Finance, vol. 42, no. 3, pp. 483–510, 1987. [Online]. Available: https://doi.org/10.1111/j.1540-6261.1987.tb04565.x

[2] D. Easley and M. O'Hara, "Information and the Cost of Capital," The Journal of Finance, vol. 59, no. 4, pp. 1553–1583, 2004. [Online]. Available: https://doi.org/10.1111/j.1540-6261.2004.00672.x

[3] R. A. Lambert, C. Leuz and R. E. Verrecchia, "Accounting Information, Disclosure and the Cost of Capital," Journal of Accounting Research, vol. 45, no. 2, pp. 385–420, 2007.

[4] Y. Amihud, "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects," Journal of Financial Markets, vol. 5, no. 1, pp. 31–56, 2002.

[5] A. Bodnaruk and P. Östberg, "The Shareholder Base and Payout Policy," Journal of Financial and Quantitative Analysis, vol. 48, no. 6, pp. 2087–2122, 2013.

7. Acknowledgment

I would like to convey my heartfelt appreciation to Professor Benedetta Montanaro of the Politecnico di Milano for her insightful guidance, rigorous feedback and support in shaping the econometric framework of this thesis. My sincere thanks also go to Professor Andriy Bodnaruk of the University of Illinois Chicago, who initiated this research project and supervised its progression.
The joint academic pathway between the two institutions, while at times requiring careful coordination and adaptation, has proven to be an immensely enriching experience both intellectually and personally. I am profoundly grateful to the dual-degree program between the Politecnico di Milano and the University of Illinois Chicago for making possible such a distinctive and transformative academic journey.