U.PORTO

MASTER
FINANCE

Price clustering in ESG stocks: Do investor preferences shape market microstructure?

Vitor Hugo Palhares Dias

M

2026



U.PORTO

FEP FACULDADE DE ECONOMIA E GESTÃO UNIVERSIDADE DO PORTO

Price clustering in ESG stocks: Do investor preferences shape market microstructure?

Vitor Hugo Palhares Dias

Dissertation

Master in Finance

Supervised by

Júlio Fernando Seara Sequeira da Mota Lobão

2026

Acknowledgments

Com a entrega e defesa da presente dissertação, encerra-se uma importante etapa da minha vida académica e pessoal. Um percurso que começou e termina na Faculdade de Economia do Porto, instituição que me abriu as portas para uma experiência memorável, feita de várias batalhas e muitas conquistas e que, acima de tudo, culminou numa evolução pessoal e intelectual pela qual só posso estar profundamente agradecido.

Desde a licenciatura até ao mestrado, foram muitas as pessoas que se cruzaram no meu caminho. Colegas e professores que, de diferentes formas, contribuíram para a minha evolução ao longo destes cinco anos. A todos eles quero deixar, nesta espécie de despedida da vida académica, o meu sincero agradecimento por terem feito parte deste percurso. Em especial, quero agradecer ao meu colega Daniel Maia, que esteve presente ao longo de toda esta caminhada, acompanhando-me desde o primeiro dia da Licenciatura em Economia até ao último dia do Mestrado em Finanças.

Quero também agradecer, em particular, ao Professor Júlio Lobão, meu professor e orientador da presente dissertação, pela sugestão do tema, pela disponibilidade e por todo o apoio prestado ao longo da realização deste trabalho.

Fora da esfera académica, quero deixar um agradecimento especial à minha família e à minha namorada, que foram, sem dúvida, o meu maior apoio ao longo de toda esta jornada. Estiveram sempre presentes e deram-me a força e o apoio necessários para ultrapassar as adversidades que foram surgindo pelo caminho.

A todos, o meu sincero obrigado.

Resumo

O *price clustering* reflete uma concentração sistemática dos preços em determinados níveis. Este fenómeno é relevante para o debate sobre a eficiência dos mercados, pois evidencia padrões de formação de preços que diferem de uma distribuição uniforme entre os diferentes incrementos disponíveis. Esta dissertação examina se a classificação ambiental, social e de governação (ESG) está associada ao *price clustering* nos mercados acionistas dos EUA. Utilizando dados diários obtidos através do LSEG Workspace para 1.000 empresas cotadas nos EUA entre janeiro de 2020 e dezembro de 2025, a análise compara 500 empresas pertencentes ao quartil superior da pontuação ESG com 500 empresas pertencentes ao quartil inferior. Os resultados confirmam a presença de *price clustering* e mostram que as empresas com elevada pontuação ESG apresentam menor *price clustering* em comparações não condicionadas. Contudo, após controlar pelo preço, taxa de rotação, volatilidade do mercado e da própria ação, dimensão da empresa e valor transacionado, o coeficiente ESG reduz-se substancialmente em magnitude e torna-se estatisticamente insignificante. Embora o coeficiente permaneça negativo em todas as especificações de robustez, a sua significância estatística varia entre especificações. De um modo geral, os resultados sugerem que parte da diferença observada no *price clustering* entre empresas com elevada e baixa pontuação ESG reflete diferenças nas características das empresas e das transações, e não apenas a classificação ESG. Os resultados obtidos, ao evidenciarem a persistência de padrões sistemáticos na formação dos preços, têm implicações para o debate sobre a eficiência dos mercados e também implicações práticas para reguladores e investidores interessados na forma como as características das empresas e dos participantes no mercado se relacionam com a formação de preços.

Palavras-chave: ESG; *price clustering*; microestrutura de mercado; preferências dos investidores; resolução de preços.

Abstract

Price clustering reflects a systematic concentration of prices at particular levels. This phenomenon is relevant to the debate on market efficiency, as it reveals price formation patterns that differ from a uniform distribution across the available price increments. This dissertation examines whether environmental, social, and governance (ESG) classification is associated with price clustering in U.S. equity markets. Using daily data from LSEG Workspace for 1,000 U.S.-listed firms between January 2020 and December 2025, the analysis compares 500 firms in the highest ESG-score quartile with 500 firms in the lowest quartile. The results confirm the presence of price clustering and show that high-ESG firms exhibit lower clustering in unconditional comparisons. However, after controlling for price, turnover, market and stock volatility, firm size, and trade value, the ESG coefficient becomes substantially smaller and statistically insignificant. Although the coefficient remains negative across all robustness specifications, its statistical significance varies across specifications. Overall, the findings suggest that part of the observed difference in price clustering between high-ESG and low-ESG firms reflects differences in firm and trading characteristics rather than ESG classification alone. The results obtained, by revealing the persistence of systematic patterns in price formation, have implications for the debate on market efficiency and also practical implications for regulators and investors interested in how firm characteristics and the characteristics of market participants relate to price formation.

Keywords: ESG; price clustering; market microstructure; investor preferences; price resolution.

Table of Contents

1 Introduction 1

2 Literature Review 4

2.1 Price clustering in equity markets 4

2.1.1 Definition, persistence, and market-design explanations 4

2.1.2 The Attraction Hypothesis 5

2.1.3 The Negotiation Hypothesis 6

2.1.4 Price Resolution Hypothesis and the determinants of clustering 6

2.2 ESG investors and preference heterogeneity 8

2.3 Why may clustering differ in ESG versus non-ESG stocks? 9

3 Data and Methodology 11

3.1 Data sources and sample construction 11

3.2 Measurement of price clustering 12

3.3 Explanatory variables 13

3.4 Econometric specification 14

3.5 Robustness strategy 15

4 Empirical Results 16

4.1 Descriptive statistics by ESG group 16

4.2 Correlation analysis 17

4.3 Last-digit distribution, HHI, and uniformity tests 17

4.4 Probit estimates 19

4.5 Average marginal effects 21

4.6 Robustness and diagnostic checks 22

4.6.1 Multicollinearity diagnostics 22

4.6.2 Year and sector fixed effects 23

4.6.3 Alternative control specifications 23

4.6.4 Robustness assessment 24
4.7 Interpretation of the Main Findings 24
5 Conclusion 26
References 29

List of Tables

Table 1. Construction and composition of the sample 12
Table 2. Descriptive statistics for the regression sample 16
Table 3. Correlation matrix for the explanatory variables 17
Table 4. Distribution of final digits and measures of price clustering 18
Table 5. Probit estimates of the determinants of price clustering 19
Table 6. Average marginal effects from the full Probit models 21
Table 7. Variance inflation factors and tolerance 23
Table 8. Robustness of the ESG association 24

1 Introduction

Price clustering is a persistent feature of financial markets in which transaction prices tend to concentrate at particular price points rather than being uniformly distributed across all available pricing increments. Harris (1991) shows that this phenomenon can occur because market participants use coarser price resolutions than those that are permitted by the market, reducing negotiation and processing costs. Price clustering may therefore contain information about how investors process information, negotiate prices, and interact with the trading environment. This phenomenon has remained relevant even after the introduction of finer pricing grids. Ikenberry and Weston (2008) document that clustering persisted in U.S. equity markets following decimalization, suggesting that the choice of particular price points cannot be explained only by the minimum tick size. More recently, Lobão et al. (2024) also document substantial clustering in Scandinavian equity markets, reinforcing the view that preferred price points remain an important feature of modern market microstructure.

The literature has proposed several complementary explanations for this behavior. The Attraction Hypothesis emphasizes investors' preference for salient numbers, particularly prices ending in zero or five, while the Negotiation and Price Resolution Hypotheses relate clustering to the costs of precise pricing and to the degree of uncertainty surrounding an asset's fundamental value. Harris (1991) argues that price level, liquidity, and information conditions can influence the effective price resolution adopted by traders, while subsequent studies show that variables such as volatility, trading activity, and firm characteristics are also associated with clustering. Therefore, price clustering is not merely a mechanical property of the trading system, it can reflect differences in the characteristics of securities and in the behavior of the investors who trade them.

This behavioral dimension provides a natural connection with the growing literature on environmental, social, and governance (ESG) investing. ESG-oriented investors may differ from conventional investors in their motivations, investment horizons, and responses to financial and non-financial information. Bollen (2007) finds that flows into socially responsible investment funds are less sensitive to short-term performance, suggesting a more persistent investor base. Riedl and Smeets (2017) show that investors may be willing to accept financial trade-offs in order to align their portfolios with social preferences. Hartzmark and Sussman (2019) provide evidence that sustainability information can generate systematic reallocations of capital. These findings suggest that ESG preferences may affect not only

portfolio selection but also the composition and trading behavior of the investor base. More directly, Baig et al. (2022) show that differences in investor ownership can be associated with differences in price clustering, finding lower clustering among stocks with greater mutual-fund ownership. Taken together, this evidence suggests the possibility that ESG and non-ESG stocks may exhibit different microstructure patterns.

Despite the extensive literature on the determinants of price clustering and the expanding literature on ESG-related investor behavior, there is limited direct evidence on whether ESG classification is associated with differences in price clustering. This dissertation addresses this gap by testing whether high-ESG and low-ESG U.S. stocks exhibit different clustering patterns. More precisely, the analysis asks whether high-ESG firms are associated with lower price clustering than low-ESG firms and whether any observed difference persists after controlling for established microstructure determinants such as stock price, turnover, market and stock volatility, firm size, and trade value. The distinction between the unconditional and conditional comparisons is relevant because ESG classification may itself be associated with firm characteristics that also influence price resolution.

The empirical analysis uses daily data from LSEG Workspace for the period from January 1, 2020, to December 31, 2025. The sample consists of 1,000 U.S. firms, with 500 firms classified in the highest ESG-score quartile and 500 firms classified in the lowest quartile. Firms in the two middle quartiles are excluded from the main comparison. Price clustering is examined using the final decimal digit of daily closing prices. Two binary measures are considered: PC0, which identifies prices ending in zero, and PC05, which identifies prices ending in either zero or five. The analysis also examines the full distribution of terminal digits using frequency measures, the Herfindahl-Hirschman Index, and chi-square goodness-of-fit tests. The multivariate analysis is based on pooled Probit models with firm-clustered standard errors and controls for the main determinants of price resolution identified in the literature.

The results show that price clustering is present in the U.S. sample and that prices ending in zero and five occur more frequently than would be expected under a uniform distribution. In unconditional comparisons, high-ESG firms exhibit lower clustering than low-ESG firms, both in terms of PC0 and PC05 and in the overall concentration of terminal digits. The benchmark Probit models also produce a negative and statistically significant coefficient on the ESG indicator. However, once the full set of observable firm and trading characteristics is included, the magnitude of the ESG coefficient decreases substantially and becomes

statistically insignificant. The coefficient remains negative across all robustness specifications, although its statistical significance depends on the treatment of firm size and on the inclusion of year and sector effects. These results suggest that part of the unconditional difference in clustering between high-ESG and low-ESG firms is associated with differences in their observable characteristics rather than with ESG classification alone. Therefore, the findings are interpreted as evidence of an association rather than as a causal effect of ESG preferences on price clustering.

The remainder of the dissertation is organized as follows. Section 2 reviews the literature on price clustering and ESG-related investor behavior and develops the theoretical motivation for the empirical analysis. Section 3 describes the data, sample construction, variables, and econometric methodology. Section 4 presents the empirical results and robustness tests. Section 5 concludes and discusses the main implications of the findings.

2 Literature Review

This literature review is organized in two complementary parts. First, the summarization of market microstructure evidence on price clustering, focusing on how and why prices tend to concentrate at salient levels. Second, the ESG investing literature that documents behavioral differences between ESG-oriented and conventional investors. Bringing these parts together provides the conceptual basis for expecting potential differences in clustering between ESG and non-ESG stocks and motivates the hypotheses and empirical approach of this study.

2.1 Price clustering in equity markets

2.1.1 Definition, persistence, and market-design explanations

A well-established empirical regularity in financial markets is that prices are not evenly distributed across all possible price increments. Instead, transaction prices and quotes tend to concentrate at some "preferred" or salient values, often round numbers, rather than being uniformly distributed across the pricing grid. This phenomenon is known as price clustering (Harris, 1991). Early empirical evidence documents clustering in stock prices and shows that the distribution of prices is not the one we would expect if all price points were equally likely (Niederhoffer, 1965). Related evidence suggests that trading and market-making behavior can interact with discrete price grids in a way that produces predictable patterns around certain price levels (Niederhoffer & Osborne, 1966).

Microstructure research offers several explanations for the occurrence of clustering. One foundational view is that clustering reflects discreteness and trading frictions: even when markets allow very precise pricing increments, traders may prefer to use wider increments to reduce negotiation complexity, cognitive effort, or order-handling costs. In this framework, the effective "resolution" used by market participants can be larger than the minimum tick size set by the exchange, leading to concentrations of prices at salient points (Harris, 1991). This perspective implies that clustering is not purely mechanical, it can vary with liquidity, volatility, and other stock-level characteristics that shape the costs and benefits of precise pricing (Harris, 1991).

A second mechanism that, according to the literature, explains this phenomenon is the strategic behavior and incentives within the trading system. Christie and Schultz (1994) document a striking pattern in NASDAQ quoting, market makers avoiding certain quote fractions. The authors argue that the absence of "odd-eighth" quotes is difficult to explain

purely by innocent costs, highlighting the role of incentives and strategic quoting in shaping observed price grids (Christie & Schultz, 1994). The evidence therefore suggests that institutional features and strategic behavior can shape clustering patterns, independently of any purely psychological explanation (Christie & Schultz, 1994).

U.S. decimalization provides an important benchmark for modern equity markets. If clustering were only a mechanical consequence of coarse pricing increments, one would expect it to weaken substantially when the pricing grid becomes finer. Yet evidence indicates that clustering remains meaningful after decimalization, with trades still disproportionately occurring at certain preferred increments (Ikenberry & Weston, 2008).

This persistence after decimalization is important for the present setting because U.S. equities trade on a fine minimum pricing grid. Ikenberry and Weston (2008) show that decimalization changed the available grid without eliminating traders' preference for particular increments. Similarly, Hameed and Terry (1998) find that changes in tick size affect, but do not mechanically determine, clustering. Consequently, the existence of a one-cent minimum tick does not imply that all admissible terminal digits should be used with equal frequency; market participants can still choose a coarser effective price grid.

Price clustering is also relevant to the debate on market efficiency. Under the Efficient Market Hypothesis (Fama, 1970), asset prices reflect the available information. The persistence of systematic price patterns does not necessarily imply market inefficiency, but it indicates that the price formation process may be influenced by behavioral preferences, trading frictions, and market microstructure mechanisms.

2.1.2 The Attraction Hypothesis

A behavioral explanation for this phenomenon is the Attraction Hypothesis, which attributes clustering to a psychological preference for specific numbers and price points. Goodhart and Curcio (1991) argue that some terminal digits are more salient and easier for the human mind to process than others, with round numbers, specifically zero and five, acting as natural reference points. In equity markets, this mechanism predicts a disproportionate use of prices ending in zero and, on a smaller scale, five. Evidence from the Tokyo Stock Exchange shows that zero and five remain the most popular endings even in a setting without market makers or direct negotiation between traders, supporting an explanation consistent with the Attraction Hypothesis rather than one derived from collusion or bilateral negotiation (Aşçıoğlu et al., 2007).

If terminal digits were distributed in a uniform manner, each of the ten digits from zero to nine would be expected to appear approximately 10% of the time, while zero and five taken together would appear approximately 20% of the time. Under the attraction mechanism, zero should be the most frequent terminal digit and five should also be overrepresented relative to a uniform benchmark (Goodhart & Curcio, 1991; Lobão et al., 2024). These predictions provide the theoretical basis for examining the full last-digit distribution and for using PC0 and PC05 as alternative measures of price clustering.

2.1.3 The Negotiation Hypothesis

The Negotiation Hypothesis, developed by Harris (1991), states that traders use coarser price increments because round prices reduce transaction costs and simplify the negotiation and processing of transactions. Limiting the set of prices considered can accelerate agreement and reduce the number of economically minor counter offers. The relative cost of rounding is expected to decline as the stock price increases, since, for a higher-priced stock, a fixed monetary increment represents a smaller proportion of the stock value. The hypothesis therefore predicts a positive association between the price level and price clustering (Harris, 1991; Hameed & Terry, 1998; Hu et al., 2017). Based on this hypothesis, stock price is included in both the benchmark and full empirical specifications, with an expected positive association with clustering.

The tick size is also relevant for the negotiation mechanism, since it defines the minimum increment in price that is permitted. Prior evidence suggests that changes in tick size can change the clustering pattern and the effective price grid that traders are allowed to use (Hameed & Terry, 1998). However, the minimum tick in this study is fixed at one cent throughout the sample and thus it provides no cross-sectional or time-series variation in clustering. From the perspective of understanding the available pricing grid, tick size is still a relevant concept, but it is not included as a regressor in the empirical models due to its lack of variation.

2.1.4 Price Resolution Hypothesis and the determinants of clustering

A complementary explanation is the Price Resolution Hypothesis of Ball et al. (1985). The main idea is that the greater the uncertainty about the underlying value of the asset, or the weaker the information required to price the asset precisely, the coarser the pricing grid that market participants use. As uncertainty declines, and information becomes more abundant, traders should be able to distinguish between finer price points, and therefore the need to

round prices should decrease. This framework relates price clustering directly to observable characteristics of the stock and its trading environment (Ball et al., 1985; Harris, 1991).

Volatility is commonly used as a proxy for valuation uncertainty. Higher individual-stock volatility should increase clustering, since a precise terminal digit is less economically meaningful when the underlying value is changing substantially (Harris, 1991; Aitken et al., 1996; He & Wu, 2006). The same logic applies to market-wide uncertainty. Higher aggregate uncertainty can impair precise valuation and encourage the use of rounder prices. For example, Hu et al. (2017) find a greater clustering for firms with higher return volatility and when market volatility is higher. Both stock-specific volatility and lagged market volatility are therefore included in the full model, with positive expected coefficients.

The literature also links liquidity and trading activity with price clustering. Liquid securities generally allow for faster incorporation of information and a more continuous interaction between buyers and sellers, which should support a finer degree of price resolution. Thus, Harris (1991) predicts that the higher the frequency of transactions, the lower the clustering. Liu and Witte (2013) find that markets with more liquidity have less clustering. In this dissertation, turnover is used as the relative liquidity proxy, leading to an expected negative relation between turnover and the probability of observing a clustered price.

Clustering is also expected to be negatively related to firm size. Analysts and market participants tend to follow large firms more intensively and they tend to have a richer information environment. This reduces uncertainty about fundamental value and allows for finer price resolution (Harris, 1991). Trade size, in contrast, is expected to be positively related to clustering. Easley and O'Hara (1987) link larger trades with informed trading, while the price clustering literature suggests that large transactions may be related to coarser negotiated prices and, consequently, higher clustering (Lobão et al., 2024). Although prior studies typically refer to trade size, this dissertation uses daily trade value, calculated from daily closing price and trading volume, as an aggregate proxy for this trading variable. Daily trade value and firm size are included in the full specification as additional determinants of price resolution.

Clustering is persistent in equity markets and can be driven by market design, trading costs, strategic incentives, and behavioral mechanisms (Harris, 1991; Christie & Schultz, 1994; Ikenberry & Weston, 2008), as the clustering literature, taken together, shows.

The literature therefore guides the expected directional association for the empirical design of this dissertation. Price, stock volatility, market-wide volatility and daily trade value are expected to be positively associated with price clustering. Turnover and firm size are expected to be negatively associated with clustering.

2.2 ESG investors and preference heterogeneity

A second literature motivates why ESG and non-ESG equities may differ in their trading environment. The central theme is that ESG investing can reflect different investor preferences, including non-financial motives. Riedl and Smeets (2017) show that investors may hold social responsible funds motivated by their values and are able to accept financial trade-offs to align their portfolios with ethical preferences (Riedl & Smeets, 2017). This perspective implies that the demand from ESG-oriented investors can be less purely based on return chasing and potentially more stable, with important implications for trading patterns (Riedl & Smeets, 2017).

Consistent with this idea, evidence from mutual funds indicates that responsible investors can have a different behavior from conventional investors. Bollen (2007) finds that SRI-related fund flows can be less sensitive to short-term performance, suggesting a more "loyal" or sticky investor base (Bollen, 2007). Even though the empirical setting of this dissertation is equity prices rather than mutual fund flows, this is directly relevant as a behavioral foundation. If ESG-oriented investors are less reactive and more long-term oriented, they may trade less frequently and place orders in a different manner (Bollen, 2007).

At the same time, ESG demand can be highly affected by sustainability signals and labels. Hartzmark and Sussman (2019) provide evidence that sustainability rankings affect capital flows, indicating that investors pay attention to ESG information and that reclassification into sustainability categories/levels can drive portfolio reallocations (Hartzmark & Sussman, 2019). It shows that ESG preferences are strong enough to move capital in systematic ways, which may affect liquidity and the microstructure environment of ESG stocks (Hartzmark & Sussman, 2019).

More broadly, research on investing based on norms illustrates how values can segment markets and reshape ownership patterns. Hong and Kacperczyk (2009) argue that social norms can lead some investors to avoid "sin stocks," affecting who holds these assets and how they are priced (Hong & Kacperczyk, 2009). Even though sin stocks are not ESG, the

study demonstrates that values-based investing can segment the investor base, which can affect trading patterns and market outcomes (Hong & Kacperczyk, 2009).

Together, these studies suggest that ESG preferences can affect investment decisions, how investors operate in the market, and, potentially, the market microstructure.

2.3 Why may clustering differ in ESG versus non-ESG stocks?

The previous evidence suggests that ESG and non-ESG stocks may differ not only in firm characteristics but also in the composition and behavior of their investor bases. These differences may affect the mechanisms underlying price clustering.

The microstructure evidence indicates that clustering depends not only on tick size, but also on how traders place orders, how liquidity is supplied, and how salient price points are used in practice (Harris, 1991; Ikenberry & Weston, 2008). At the same time, the ESG literature suggests that ESG-oriented investors may differ in motives, attention, and responsiveness, which can shape turnover and trading strategies (Riedl & Smeets, 2017; Bollen, 2007; Hartzmark & Sussman, 2019).

Behavioral mechanisms provide an additional link between investor preferences and price clustering. Bhattacharya et al. (2012) document systematic buy and sell imbalances around round numbers, consistent with traders using salient price points as anchors. Kavajecz and Odders-White (2004) also show that liquidity provision can concentrate around widely observed technical price levels. Differences in the use of such reference points across investor groups could therefore translate into different levels of price clustering.

If ESG-oriented investors are more long-term oriented and less driven by short-term signals, they may trade less frequently and rely less on simple heuristics such as round-number anchoring, potentially reducing clustering (Riedl & Smeets, 2017; Bollen, 2007). Ownership composition may also affect clustering through liquidity provision and trading costs, since ESG screening can influence who holds a stock (Harris, 1991; Hong & Kacperczyk, 2009). Differences in order placement may further contribute if buying and selling pressure is distributed differently around salient price levels (Bhattacharya et al., 2012; Kavajecz & Odders-White, 2004).

Baig et al. (2022) show that stocks with higher mutual-fund ownership exhibit less price clustering, directly demonstrating that the characteristics of the investor base can affect the degree of price clustering. Accordingly, even if ESG status itself is not the ultimate causal

mechanism, a comparison of ESG and non-ESG firms is still economically meaningful, since ESG classification may be associated with different ownership structures, trading intensity, information environments, and order-placement behavior, all of which can alter the effective price grid selected by market participants.

ESG and non-ESG firms also differ in characteristics that are already known to influence clustering, including firm size, liquidity, volatility, price level and trading value (Harris, 1991; Hu et al., 2017; Lobão et al., 2024). The multivariate analysis therefore controls for these characteristics to assess whether the ESG association remains after accounting for established determinants of price clustering.

The theoretical prediction of the ESG indicator is thus directional but should be interpreted with caution. If high-ESG stocks have a more stable, long-term or institutionally dominated investor base and rely less on round-number heuristics, they should exhibit less price clustering than low-ESG stocks (Bollen, 2007; Riedl & Smeets, 2017; Baig et al., 2022). However, since ESG status is also associated with firm and trading characteristics, the existing literature does not suggest that this negative relationship should remain statistically significant after controlling for the standard determinants of price clustering. A reduction in the ESG coefficient after adding these controls would therefore suggest that part of the initial difference in clustering between high-ESG and low-ESG firms is explained by differences in their firm and trading characteristics.

The clustering literature provides strong evidence on the existence and determinants of clustering in equities, while the ESG literature documents meaningful investor heterogeneity and demand driven by preferences. However, there is limited direct evidence on whether ESG and non-ESG equity stocks exhibit systematically different clustering intensity. This dissertation aims to provide a response to this gap by testing whether ESG classification is associated with distinct clustering outcomes and, more broadly, whether ESG-oriented investors contribute to different microstructure patterns relative to conventional stocks.

3 Data and Methodology

3.1 Data sources and sample construction

The methodology follows the general sequence used by Lobão, Pacheco, and Carvalho (2024): first, the distribution of price-ending digits is evaluated through frequencies, the Herfindahl-Hirschman Index (HHI), and a chi-square goodness-of-fit test, second, Probit models are estimated to identify the determinants of clustering. The present study extends that framework by introducing an ESG indicator as the variable of primary interest and by comparing high-ESG and low-ESG firms within the United States equity market.

Daily market data were obtained from LSEG Workspace for U.S.-listed equities over the period from January 1, 2020, to December 31, 2025. The initial universe consisted of U.S. companies with a market capitalization exceeding USD 1 billion and an available Refinitiv ESG Score. Following the quartile-based classification approach adopted by Lins, Servaes, and Tamayo (2017), firms were ranked by ESG score, with those in the highest quartile classified as ESG and those in the lowest quartile classified as non-ESG. Firms in the two middle quartiles were excluded from the principal comparison, producing a final sample of 500 ESG and 500 non-ESG firms. The variables used in the empirical analysis include daily closing prices, trading volume, shares outstanding, market capitalization, and the S&P 500 index.

The ESG classification used is time-invariant for each firm and is based on the score available as of May 2025, date of extraction. This approach treats ESG classification as a persistent firm-level characteristic, not allowing firms to move between ESG groups over the sample period, consistent with the broader use of persistent sustainability classifications in the literature (Eccles et al., 2014). The results must therefore be interpreted as a comparison between firms ranked in the top and bottom ESG quartiles, classified as ESG and non-ESG firms, respectively, rather than as evidence that annual changes in ESG performance caused changes in price clustering.

A balanced firm-date grid initially contained 1,566,000 observations. Observations without valid closing prices were excluded from the digit distribution analysis. The regression sample was then restricted to complete observations for all dependent and explanatory variables. In particular, the 60-day rolling stock volatility measure and the one-day lag of market volatility necessarily reduce the usable sample at the beginning of firm histories. Non-trading

observations and observations for which logarithmic transformations were undefined were also excluded.

Table 1 summarizes the sample construction and composition across the initial panel, the digit-analysis sample, and the regression sample.

Table 1. Construction and composition of the sample

Metric	Value	% of initial panel
Full panel observations	**1,566,000**	**100.00%**
Valid closing-price observations used for digit analysis	**1,474,459**	**94.15%**
Regression sample observations	**1,305,519**	**83.37%**
Regression sample firms	**1,000**	
ESG firms	**500**	
Non-ESG firms	**500**	
Initial panel dates	**1,566**	
Regression sample trading dates	**1,395**	
Average observations per firm (regression)	**1,305.52**	
Sample start	**2020-01-01**	
Sample end	**2025-12-31**	

Notes: *The HHI and digit-frequency analysis uses all observations with a valid closing price, whereas the regression sample uses complete cases for all model variables. Percentages are calculated relative to the initial firm-date panel.*

The table above shows that the digit analysis retains 94.15% of the initial panel and that the complete regression sample retains 83.37%. The reduction is material but not unexpected given the use of rolling volatility and lagged market variables. The number of firms remains at 1,000, which limits concerns that the filter removed entire companies.

3.2 Measurement of price clustering

Price clustering is measured from the final decimal digit of each daily closing price. Two binary dependent variables are used, PC0, that equals one when the final digit is zero and zero otherwise, and PC05, that equals one when the final digit is either zero or five and zero otherwise. PC0 is the principal definition, it is consistent with studies that treat zero as the strongest round-number focal point, PC05 provides an alternative definition that captures the two most salient decimal endings (Aitken et al., 1996; Ikenberry & Weston, 2008; Lobão et al., 2024).

The Herfindahl-Hirschman Index, HHI, was used to summarize the degree of concentration across all ten possible final digits, following Ikenberry and Weston's (2008) adaptation of the

index to the analysis of price clustering. Let f_i denote the proportion of valid closing prices ending in digit i, where i ranges from zero to nine.

The HHI is calculated as:

$$HHI = \sum_{i=0}^{9} f_i^2 \tag{3.1}$$

Following Ikenberry and Weston (2008), and consistent with Lobão, Pacheco, and Carvalho (2024), the HHI is used to measure the concentration of closing price final digits. When f_i denotes the relative frequency of prices ending in digit i, expressed as a proportion, a uniform distribution across the ten possible digits produces an HHI of 0.10. Progressively larger values indicate greater concentration in a smaller number of terminal digits, while complete concentration in a single digit produces the maximum value of one.

In addition to HHI, which provides a descriptive measure of the overall concentration of terminal digits, the chi-square goodness-of-fit test was employed to assess whether the observed digit distribution differs statistically from the frequencies expected under a uniform distribution (Lobão, Pacheco, & Carvalho, 2024). Given the very large number of observations, the chi-square test is not interpreted in isolation, since even economically small departures from the null hypothesis may generate very low p-values in large samples (Lin, M., Lucas Jr, H. C., & Shmueli, G., 2013). The chi-square results are evaluated jointly with the HHI, the observed digit frequencies, and the magnitude of the departures from the uniform distribution.

The statistic is calculated as:

$$\chi^2 = \sum_{i=0}^{9} \frac{(O_i - E_i)^2}{E_i} \tag{3.2}$$

where O_i and E_i are the observed and expected counts for digit i. Under the null hypothesis, the statistic follows a chi-square distribution with nine degrees of freedom. This is the same broad univariate testing strategy applied by Lobão et al. (2024).

3.3 Explanatory variables

The explanatory variables are selected from the negotiation and price resolution explanations of clustering.

ESG Dummy is the main variable of interest and equals one for firms in the top ESG-score quartile and zero for firms in the bottom quartile. Ln (Price) is the natural logarithm of the daily closing price. Ln (Turnover) is the natural logarithm of daily volume scaled by shares outstanding and serves as a relative liquidity measure. Ln (Market Volatility t-1) is the logarithm of the absolute S&P 500 return on the preceding trading day. The lag makes the measure predetermined relative to the current closing-price ending. Ln (Stock Volatility 60D) is the logarithm of the rolling standard deviation of daily stock returns over the previous 60 trading days. Ln (Firm Size) is based on average market capitalization over the sample period. Trade Value is calculated as the daily closing price multiplied by daily trading volume, and Ln (Trade Value) is the natural logarithm of this measure.

The expected signs follow prior microstructure research. Price, stock volatility, market volatility, and daily trade value are expected to increase clustering, while turnover and firm size are expected to reduce it. Tick size was not included as an explanatory variable because it was constant at $0.01 across the sample and therefore provided no cross-sectional or time-series variation. However, tick size remains relevant to the measurement of price clustering, as it determines the minimum permitted price increment.

3.4 Econometric specification

Because PC0 and PC05 are binary variables, the determinants of clustering are estimated using pooled Probit models, in which the probability of observing price clustering is linked to the explanatory variables through the cumulative standard normal distribution, Φ. The benchmark specification includes ESG classification, price, and turnover:

$$\Pr(PC_{it} = 1 \mid X_{it}) = \Phi(\alpha + \beta_1 ESG_i + \beta_2 \, ln\, Price_{it} + \beta_3 \, ln\, Turnover_{it}) \quad (3.3)$$

The full specification adds market volatility, stock-specific volatility, firm size, and daily trade value:

$$\begin{aligned}\Pr(PC_{it} = 1 \mid X_{it}) = \Phi(&\alpha + \beta_1 ESG_i + \beta_2 \ln Price_{it} + \beta_3 \ln Turnover_{it} \\ &+ \beta_4 \ln MarketVol_{t-1} + \beta_5 \ln StockVol60D_{it} \\ &+ \beta_6 \ln FirmSize_i + \beta_7 \ln TradeValue_{it})\end{aligned} \quad (3.4)$$

where Φ denotes the cumulative standard normal distribution. Each model is estimated twice, first using PC0 and then using PC05. The comparison between the benchmark and full specifications is central to the research question, it shows whether an initial ESG

association remains after observable differences in size, volatility, liquidity, and trading activity are incorporated.

All Probit specifications are estimated with standard errors clustered at the firm level (given the repeated daily observations for each firm). Clustering the standard errors ensures that inference is not compromised by within-firm dependence of repeated observations, a common issue for financial panel data (Petersen, 2009; Cameron & Miller, 2015). Also included in the robustness analysis are year and sector indicators, which control for common time shocks and systematic heterogeneity at the sector level (Gormley & Matsa, 2014).

Probit coefficients show direction and statistical significance of association but are not direct percentage-point effects on probability. Therefore, average marginal effects (AMEs) are calculated for complete and robustness specifications. The marginal effect for each observation is evaluated at the observation's covariate values and then averaged across the estimation sample. For the ESG indicator, the AME is the average discrete change in predicted probability when the dummy changes from zero to one, with the other variables held at their observed values. The AME inference reported uses the same firm-clustered covariance structure as the corresponding Probit model. AMEs are preferred because they offer a more transparent probability-scale interpretation of the estimates from nonlinear models (Williams, 2012; Cameron & Miller, 2015).

3.5 Robustness strategy

The purpose of the robustness and diagnostic checks is to test if the central ESG finding is sensitive to alternative definitions, observable heterogeneity and overlapping explanatory information. First, PC05 offers the alternative definition of price clustering, by capturing prices ending in either zero or five. Second, we augment the full Probit specifications with year and sector fixed effects to control for common annual shocks and systematic sector-level differences that might otherwise affect the conditional comparison (Gormley & Matsa, 2014). Third, multicollinearity is checked by variance inflation factor (VIF) and tolerance value. VIF thresholds should not be used as mechanical deletion rules (O'Brien, 2007), the diagnostics are interpreted in conjunction with the correlation matrix and with alternative specifications without Firm Size and Trade Value separately. Finally, we report the cluster-robust AMEs of each robustness specification to compare the economic magnitude of the ESG association on the probability scale (Williams, 2012).

4 Empirical Results

4.1 Descriptive statistics by ESG group

Table 2 presents descriptive statistics for the regression sample and compares the means of the main variables across ESG and non-ESG observations, including tests of the statistical significance of the mean differences.

Table 2. Descriptive statistics for the regression sample

Variable	Full sample mean	Full sample SD	ESG mean	Non-ESG mean	Difference
PC0	**0.1204**	**0.3254**	**0.1147**	**0.1268**	**-0.0121*****
PC05	**0.2257**	**0.4181**	**0.2185**	**0.2339**	**-0.0154*****
Ln (Price)	**3.7940**	**1.2338**	**4.2352**	**3.2998**	**0.9354*****
Ln (Turnover)	**-4.9646**	**1.0573**	**-4.9176**	**-5.0173**	**0.0997****
Ln (Market Volatility t-1)	**-5.3596**	**1.2593**	**-5.3518**	**-5.3682**	**0.0164*****
Ln (Stock Volatility 60D)	**-3.7312**	**0.5457**	**-3.8977**	**-3.5447**	**-0.3531*****
Ln (Firm Size)	**22.5897**	**1.7149**	**23.6170**	**21.4387**	**2.1783*****
Ln (Trade Value)	**17.5445**	**2.0626**	**18.6121**	**16.3482**	**2.2639*****

*Notes: N = 1,305,519. The ESG subsample contains 689,832 observations and the non-ESG subsample contains 615,687 observations. Difference is ESG mean minus non-ESG mean. Statistical significance of mean differences is assessed using regressions of each variable on the ESG indicator with standard errors clustered at the firm level. * p < 0.10, ** p < 0.05, *** p < 0.01.*

The unadjusted evidence is consistent with lower clustering in ESG firms. The share of prices ending in zero is 11.47% for ESG observations and 12.68% for non-ESG observations, a difference of 1.21 percentage points. Considering zero and five, the proportion is 21.85% and 23.39%. The difference is 1.54 percentage points. The differences in PC0 and PC05 are statistically significant at the 1% level. The unadjusted difference in PC0 is about 9.5% of the non-ESG mean and the difference in PC05 is about 6.6%.

The table also indicates that the two groups are not directly comparable on observable stock characteristics. ESG observations are associated with higher prices, higher turnover, substantially larger firm size, higher daily trade value and lower stock-specific volatility. These differences are statistically significant at the 1% level in all cases except turnover, which is significant at the 5% level. The difference in Ln (Firm Size) of 2.1783 is particularly large and is consistent with ESG firms being many times larger in terms of market capitalization. Such differences create a high risk of composition effects: an unconditional ESG gap in clustering may capture the characteristics of firms that score high on ESG, rather than an independent ESG-investor channel.

4.2 Correlation analysis

Table 3 presents the correlation matrix for the explanatory variables used in the regression analysis.

Table 3. Correlation matrix for the explanatory variables

Variable	ESG	Ln Price	Ln Turnover	Ln Mkt Vol	Ln Stock Vol	Ln Firm Size	Ln Trade Value
ESG	1.000	0.378	0.047	0.007	-0.323	0.634	0.548
Ln Price	0.378	1.000	-0.057	-0.025	-0.430	0.596	0.554
Ln Turnover	0.047	-0.057	1.000	0.014	0.275	0.060	0.501
Ln Mkt Vol	0.007	-0.025	0.014	1.000	0.083	0.006	-0.012
Ln Stock Vol	-0.323	-0.430	0.275	0.083	1.000	-0.377	-0.215
Ln Firm Size	0.634	0.596	0.060	0.006	-0.377	1.000	0.837
Ln Trade Value	0.548	0.554	0.501	-0.012	-0.215	0.837	1.000

Notes: Pearson correlations calculated using the complete regression sample.

The correlation matrix above shows that ESG status is strongly correlated with firm characteristics. ESG_Dummy is positively related to Firm Size (0.634) and Trade Value (0.548), and negatively related with Stock Volatility (-0.323). These relationships are consistent with the descriptive evidence and offer a statistical explanation for why a simple ESG coefficient might change when additional controls are introduced.

The largest pairwise correlation is 0.837 between Firm Size and Trade Value. This does not necessarily imply problematic multicollinearity as pairwise correlations do not capture the relationship of each explanatory variable with the full set of regressors. High collinearity can increase the variance of the coefficient estimates and their standard errors, making it more difficult to distinguish the individual contribution of closely related explanatory variables (O'Brien, 2007; Dormann et al., 2013). In addition to the correlation analysis, the Variance Inflation Factor (VIF) diagnostics and alternative specifications without the Firm Size and Trade Value variables separately are applied. Therefore, the interpretation of the ESG coefficient should distinguish whether the changes across specifications stem from the inclusion of relevant firm characteristics or from a decrease in estimation precision caused by multicollinearity.

4.3 Last-digit distribution, HHI, and uniformity tests

Table 4 presents the distribution of final price digits for the full, ESG, and non-ESG samples, together with the HHI and chi-square tests of departure from a uniform distribution.

Table 4. Distribution of final digits and measures of price clustering

Digit or statistic	Uniform benchmark	Full sample	ESG	Non-ESG
0	10.00%	12.26%	11.54%	13.06%
1	10.00%	9.65%	9.81%	9.48%
2	10.00%	9.70%	9.79%	9.61%
3	10.00%	9.48%	9.67%	9.27%
4	10.00%	9.70%	9.77%	9.62%
5	10.00%	10.56%	10.36%	10.78%
6	10.00%	9.75%	9.87%	9.62%
7	10.00%	9.48%	9.63%	9.32%
8	10.00%	9.74%	9.81%	9.66%
9	10.00%	9.68%	9.76%	9.59%
0 + 5	20.00%	22.82%	21.90%	23.83%
HHI	0.1000	0.1006	0.1003	0.1012
Chi-square	df = 9	9,565.17	2,326.62	8,340.83
p-value	-	<0.001	<0.001	<0.001
N	-	1,474,459	775,724	698,735

Notes: Digit percentages are based on observations with a valid closing price. Under uniformity, each digit has an expected frequency of 10% and digits zero and five jointly have an expected frequency of 20%. The HHI equals 0.10 under a uniform distribution.

The univariate results provide strong evidence for non-uniform distribution of the final digits of closing prices. In the full sample and in both ESG groups, zero is the most frequent final digit, followed by five. In the full sample, 12.26% of prices end in zero and 22.82% in zero or five, relative to 10% and 20% under uniform benchmarks. This ordering is consistent with the Attraction Hypothesis and with substantial evidence that zero and five are cognitively salient price points (Harris, 1991; Ikenberry & Weston, 2008; Lobão et al., 2024).

The chi-square tests reject ($p < 0.001$) the null hypothesis of uniformity for the full, ESG, and non-ESG samples. The non-ESG sample has the highest frequency of PC0, the highest combined frequency of PC05 and the highest HHI. This is consistent with more pronounced unconditional clustering of non-ESG firms. However, the economic importance of the concentration is moderate. The HHI for the full sample is 0.1006, just above the uniform benchmark of 0.1000. The chi-square test is highly sensitive to small deviations from the expected distribution when sample sizes are large (McLaren et al., 1994), so the statistical rejection should not be described as evidence of extreme concentration.

The clustering in the current U.S. sample is also weaker than what was reported by Lobão et al. (2024) for the Nordic exchanges, where market-level HHIs were between about 0.1040 and 0.1205 and zero-ending prices were much more prevalent. This is a descriptive

comparison as the markets, sample periods, currencies, price levels and institutional settings are different.

4.4 Probit estimates

Table 5 reports the estimated coefficients from the benchmark and full Probit models for both PC0 and PC05.

Table 5. Probit estimates of the determinants of price clustering

Variable	PC0 Benchmark	PC0 Full	PC05 Benchmark	PC05 Full
Intercept	-1.4598*** (0.0571)	-0.8993*** (0.0812)	-0.9961*** (0.0504)	-0.6036*** (0.0864)
ESG Dummy	-0.0636*** (0.0128)	-0.0112 (0.0108)	-0.0460*** (0.0125)	-0.0084 (0.0124)
Ln (Price)	0.0112** (0.0045)	0.0241*** (0.0065)	0.0012 (0.0036)	0.0063 (0.0068)
Ln (Turnover)	-0.0554*** (0.0117)	-0.1105*** (0.0314)	-0.0527*** (0.0111)	-0.1131*** (0.0318)
Ln (Market Volatility t-1)	-	0.0005 (0.0014)	-	0.0017 (0.0013)
Ln (Stock Volatility 60D)	-	0.0904*** (0.0155)	-	0.0706*** (0.0147)
Ln (Firm Size)	-	-0.0653*** (0.0244)	-	-0.0656*** (0.0233)
Ln (Trade Value)	-	0.0514** (0.0245)	-	0.0583** (0.0245)
Observations	1,305,519	1,305,519	1,305,519	1,305,519
LR chi-square	2,413.45	4,180.65	2,622.37	4,442.83
LR p-value	<0.001	<0.001	<0.001	<0.001
McFadden R2	0.0025	0.0044	0.0019	0.0032

*Notes: Maximum-likelihood Probit coefficients are reported with firm-clustered standard errors in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01. Clustering is performed at the firm level to account for dependence across repeated daily observations within each company (Petersen, 2009; Cameron & Miller, 2015).*

The benchmark models indicate a negative, and statistically significant, relationship between ESG classification and price clustering. The ESG coefficient is -0.0636 for PC0 and -0.0460 for PC05 and both are significant at the 1% level. When the full set of observable controls are added, the ESG coefficients remain negative but are reduced in magnitude to -0.0112 and -0.0084 and are no longer statistically significant. Therefore, the evidence in the baseline full specification does not support a statistically significant ESG association after controlling for differences in price, liquidity, volatility, firm size and daily trading value.

This change is economically informative. The descriptive tables show that ESG firms are substantially larger, less volatile and more actively traded. Thus, the benchmark ESG

coefficient seems to absorb structural features that are correlated with both ESG classification and price clustering. A conservative interpretation is that part of the lower unconditional clustering of ESG stocks is driven by characteristics of firms that receive high ESG scores rather than the ESG status by itself. This should not be taken as evidence that ESG preferences do not matter, because the model is associational and the robustness specifications below show that the statistical significance of the ESG coefficient is sensitive to the set of controls and fixed effects.

In general, the control variables conform to the Negotiation and Price Resolution Hypotheses. Turnover is negative and significant in both full models, which is consistent with the proposition that more liquid stocks incorporate information at a finer price resolution (Harris, 1991; Liu & Witte, 2013). Stock-specific volatility is positive and significant, suggesting that greater uncertainty about a firm's value encourages the use of coarser, rounder prices (Ball et al., 1985; Harris, 1991; He & Wu, 2006). Firm Size is negative and significant, consistent with the argument that larger firms are associated with greater information availability and thus lower price clustering (Harris, 1991). The coefficient on Ln (Trade Value) is positive and significant. This result is directionally consistent with the broader literature that links larger trading activity to information and price-negotiation mechanisms, although the variable used in this dissertation measures aggregate daily trade value rather than individual trade size (Easley & O'Hara, 1987; Lobão et al., 2024). These patterns resemble those documented by Lobão et al. (2024), although the present analysis deals with a different market and introduces ESG status.

Price is positive in both specifications but statistically significant only for PC0, providing partial support for the Negotiation Hypothesis, which predicts greater price clustering as stock prices increase (Harris, 1991). Market volatility is not significant in either full model, indicating that lagged aggregate market uncertainty is not significantly associated with price clustering after controlling for stock-level characteristics. Previous studies have also found mixed evidence for market and stock volatility, indicating that their effects may depend on market structure and sample design (Aitken et al., 1996; He & Wu, 2006; Narayan et al., 2011).

The likelihood-ratio tests are statistically significant in all four specifications, indicating that the explanatory variables are jointly informative relative to models containing only an intercept. The full models improve on the benchmark models in terms of McFadden R2, but

the values remain low at 0.0044 for PC0 and 0.0032 for PC05. The models identify statistically meaningful associations but explain only a small share of the daily occurrence of a particular final digit. The results should therefore not be interpreted as evidence of strong predictive performance.

4.5 Average marginal effects

Table 6 reports the average marginal effects from the full Probit models for PC0 and PC05, together with their standard errors and p-values.

Table 6. Average marginal effects from the full Probit models

Variable	PC0 AME	PC0 SE	PC0 p-value	PC05 AME	PC05 SE	PC05 p-value
ESG Dummy	-0.0022	0.0022	0.299	-0.0025	0.0037	0.497
Ln (Price)	0.0048***	0.0013	<0.001	0.0019	0.0020	0.351
Ln (Turnover)	-0.0221***	0.0063	<0.001	-0.0339***	0.0096	<0.001
Ln (Market Volatility t-1)	0.0001	0.0003	0.706	0.0005	0.0004	0.189
Ln (Stock Volatility 60D)	0.0181***	0.0031	<0.001	0.0212***	0.0044	<0.001
Ln (Firm Size)	-0.0130***	0.0049	0.008	-0.0197***	0.0070	0.005
Ln (Trade Value)	0.0103**	0.0049	0.037	0.0175**	0.0074	0.018

*Notes: AME denotes the average marginal effect over the complete estimation sample. For ESG Dummy, the reported value is the average discrete change from zero to one. Standard errors and p-values are based on firm-clustered covariance estimates. ** p < 0.05, *** p < 0.01 (Williams, 2012; Cameron & Miller, 2015).*

The ESG marginal effects are quite small and not statistically significant. Controlling for the included variables at their observed values, ESG classification is associated with a reduction in the probability that a price ends in zero by 0.22 percentage points on average, and a reduction by 0.25 percentage points in the probability that a price ends in zero or five. The corresponding 95% confidence intervals are roughly [-0.65, 0.20] percentage points for PC0 and [-0.98, 0.48] percentage points for PC05. Confidence intervals for both include zero, meaning the marginal effects of ESG are not statistically different from zero at the 5% level of significance.

For continuous variables expressed in natural logarithms, the AMEs can be expressed in terms of approximate effects of doubling by multiplying the estimated marginal effect by ln (2), since doubling the underlying variable increases its natural logarithm by ln (2). However, in nonlinear probability models such as the Probit, this calculation should be regarded as a first-order approximation, rather than an exact discrete change in predicted probability (Uberti, 2017). Under this approximation, a doubling of turnover is associated with a 1.53 percentage-point reduction in PC0 and a 2.35 percentage-point reduction in PC05. Doubling

stock volatility is associated with increases of approximately 1.25 and 1.47 percentage points, respectively. Doubling firm size is associated with a reduction of approximately 0.90 and 1.36 percentage points, respectively, while doubling trade value is associated with increases of approximately 0.71 and 1.21 percentage points, respectively. Under this doubling interpretation, the estimated changes associated with the traditional microstructure variables are larger in magnitude than the adjusted ESG marginal effect in the full models.

4.6 Robustness and diagnostic checks

The robustness analysis focuses on whether the central ESG association is stable when the model controls more explicitly for time and sector heterogeneity and when the two most strongly related explanatory variables are examined separately. All robustness Probit models retain the same 1,305,519 observations and 1,000 firms, ensuring that differences across specifications reflect changes in model specification rather than changes in the estimation sample. Standard errors are clustered at the firm level in all specifications (Petersen, 2009; Cameron & Miller, 2015).

4.6.1 Multicollinearity diagnostics

The correlation matrix identifies a relatively high correlation of 0.837 between Firm Size and Trade Value. VIF and tolerance diagnostics provide a joint assessment of how strongly each explanatory variable is related to the remaining regressors. The VIF for Trade Value is 12.60 and the corresponding tolerance is 0.079, while Firm Size has a VIF of 8.54 and a tolerance of 0.117. All remaining variables have VIFs below four. These values indicate that the main multicollinearity concern is concentrated in Firm Size and Trade Value rather than in ESG Dummy itself, whose VIF is 1.71. Consistent with O'Brien (2007), the VIF values are treated as diagnostics rather than as mechanical thresholds for deleting variables.

Table 7 presents the VIF and tolerance diagnostics for the explanatory variables included in the full specification.

Table 7. Variance inflation factors and tolerance

Variable	VIF	Tolerance
Ln (Trade Value)	12.5952	0.0794
Ln (Firm Size)	8.5371	0.1171
Ln (Turnover)	3.8817	0.2576
Ln (Price)	1.9652	0.5089
ESG Dummy	1.7078	0.5856
Ln (Stock Volatility 60D)	1.4256	0.7015
Ln (Market Volatility t-1)	1.0135	0.9866

Notes: VIF denotes the variance inflation factor and tolerance equals 1/VIF. The diagnostics are calculated using the complete regression sample. VIF thresholds are interpreted cautiously and are not used as automatic variable-deletion rules (O'Brien, 2007).

4.6.2 Year and sector fixed effects

Adding year and sector fixed effects changes the conditional ESG estimate. The ESG coefficient becomes -0.0221 for PC0 (p = 0.022) and -0.0208 for PC05 (p = 0.038), both significant at the 5% level. The corresponding cluster-robust AMEs are -0.44 and -0.62 percentage points. The negative sign is therefore preserved, but the statistical inference differs from the baseline full model once common annual shocks and sector-level heterogeneity are absorbed. This sensitivity is consistent with the general role of fixed effects in controlling for systematic group-level heterogeneity, while it does not convert the model into a causal design (Gormley & Matsa, 2014).

4.6.3 Alternative control specifications

The VIF evidence suggests two alternative specifications, omitting Firm Size and Trade Value separately. When Firm Size is removed, the ESG coefficient in absolute value becomes larger, and statistically significant for both PC0 (-0.0350, p = 0.002) and PC05 (-0.0316, p = 0.007). The associated AMEs imply reductions of approximately 0.70 and 0.95 percentage points, respectively. This result should be interpreted with caution, as Firm Size is strongly related to ESG classification in the sample and is theoretically relevant to price resolution. Larger firms are expected to be associated with greater information production and finer price resolution (Harris, 1991). The stronger ESG coefficient after excluding Firm Size is thus consistent with ESG Dummy absorbing some of the firm size related cross-sectional variation.

Conversely, when Trade Value is dropped, the ESG coefficient is small and statistically insignificant for both PC0 (-0.0068, p = 0.572) and PC05 (-0.0036, p = 0.796). The corresponding AMEs are approximately -0.14 and -0.11 percentage points, respectively. This specification therefore reinforces the baseline finding that the adjusted ESG association is

small when Firm Size is kept in the model. In general, the alternative specifications reveal that the direction of the ESG association is consistently negative, but its statistical significance depends on the treatment of Firm Size and the inclusion of year and sector effects.

Table 8 summarizes the estimated ESG coefficients and average marginal effects across the baseline and robustness specifications.

Table 8. Robustness of the ESG association

Specification	PC0 ESG coefficient (SE)	PC0 ESG AME percentage points	PC05 ESG coefficient (SE)	PC05 ESG AME percentage points
Full clustered	-0.0112 (0.0108)	-0.224 (p=0.299)	-0.0084 (0.0124)	-0.253 (p=0.497)
Year + sector FE	-0.0221** (0.0097)	-0.442** (p=0.022)	-0.0208** (0.0101)	-0.624** (p=0.038)
Without Firm Size	-0.0350*** (0.0112)	-0.700*** (p=0.002)	-0.0316*** (0.0118)	-0.947*** (p=0.007)
Without Trade Value	-0.0068 (0.0121)	-0.137 (p=0.571)	-0.0036 (0.0137)	-0.107 (p=0.796)

*Notes: All specifications contain 1,305,519 observations from 1,000 firms and use firm-clustered standard errors. AMEs for ESG Dummy are average discrete changes from zero to one and are shown in percentage points. The Year + sector FE specification includes year and TRBC Economic Sector indicators. * p < 0.10, ** p < 0.05, *** p < 0.01.*

4.6.4 Robustness assessment

The robustness evidence does not allow the conclusion that ESG is significant according to a unique specification criterion. Instead, it yields a consistent directional result: the ESG coefficient and AME are negative across all specifications. The association is statistically insignificant in the baseline full model and when Trade Value is omitted, but significant after adding year and sector fixed effects, and when Firm Size is omitted. Overall, high-ESG firms are associated with lower price clustering, although the statistical significance of this association depends on the specification. In particular, the results are sensitive to the treatment of Firm Size and to the inclusion of year and sector effects. The robustness tests therefore support a consistently negative direction of association, but not statistical significance across all specifications.

4.7 Interpretation of the Main Findings

Three findings emerge from the empirical analysis. Price clustering is present in the U.S. sample, with zero and five occurring more frequently than under a uniform distribution of terminal digits. ESG firms also exhibit lower clustering in unconditional comparisons, with lower frequencies of PC0 and PC05 and lower HHI compared to non-ESG firms. Once firm

and trading characteristics are controlled for, the negative ESG association becomes smaller and its statistical significance depends on the specification.

In the baseline full Probit models, the ESG coefficient is small and statistically insignificant when controlling for price, turnover, market volatility, stock volatility, firm size and trade value. The robustness analysis qualifies but does not overturn that finding. The coefficient remains negative for all specifications, turns out to be significant after controlling for year and sector heterogeneity, and becomes significant when Firm Size is removed. By contrast, the exclusion of Trade Value makes the ESG coefficient insignificant. Because Firm Size is strongly correlated with ESG classification and is itself a theoretically relevant determinant of price resolution, the results indicate that firm composition is an important part of the observed ESG/non-ESG difference (Harris, 1991). At the same time, the year and sector specification indicates that common temporal and sectoral heterogeneity also matters for the conditional comparison (Gormley & Matsa, 2014).

Thus, this study does not interpret the ESG coefficient as a causal effect of sustainability preferences on price formation. The results therefore indicate a consistent negative association between ESG classification and price clustering, although its statistical significance varies across specifications. Traditional microstructure variables remain central to the results: higher turnover is generally associated with less clustering, stock-specific volatility with more clustering, and firm size with less clustering, in line with the negotiation and price-resolution explanations developed in the literature (Harris, 1991; He & Wu, 2006; Lobão et al., 2024).

5 Conclusion

This dissertation examines whether ESG classification is associated with price clustering in U.S. equity markets. The analysis combines the price-clustering literature, which relates clustering to behavioral preferences, negotiation costs, liquidity, and valuation uncertainty, with evidence that ESG-oriented investors may differ from conventional investors in their preferences and trading behavior.

The results confirm that price clustering remains present in U.S. equities. Prices ending in zero and five occur more frequently than expected under a uniform distribution, although the HHI values remain close to the uniform benchmark, indicating that the overall degree of concentration is moderate. High-ESG firms also exhibit lower clustering than low-ESG firms in unconditional comparisons, with lower frequencies of PC0 and PC05 and a lower HHI.

The multivariate results show that this difference is related to observable firm and trading characteristics. The ESG coefficient is negative and statistically significant in the benchmark Probit models but becomes substantially smaller and statistically insignificant after controlling for price, turnover, market volatility, stock volatility, firm size, and trade value. The coefficient remains negative across all robustness specifications, although its statistical significance varies depending on the inclusion of year and sector effects and on the treatment of Firm Size and Trade Value. The results therefore support a consistently negative association between ESG classification and price clustering, but not a statistically robust effect across all specifications.

Turnover is negatively associated with clustering, while stock-specific volatility is positively associated with it. Firm size is associated with lower clustering and trade value with higher clustering, broadly supporting the Negotiation and Price Resolution (Harris, 1991; Ball et al., 1985) explanations discussed in the literature. Overall, the evidence suggests that part of the unconditional difference between high-ESG and low-ESG firms reflects differences in their observable characteristics rather than ESG classification alone. The findings should therefore be interpreted as associational rather than causal.

From a theoretical perspective, the persistence of price clustering has implications for the debate on market efficiency. A systematic concentration of prices at particular terminal digits is difficult to reconcile with a simple benchmark in which all permitted price increments are

used uniformly. The evidence that zero and five occur with greater frequency indicates that price formation is influenced by recurring microstructure and behavioral patterns. However, these results should not be interpreted as a direct rejection of the Efficient Market Hypothesis (Fama, 1970). Price clustering may arise from rational responses to negotiation costs, information uncertainty, liquidity conditions, or other market frictions, and therefore does not necessarily imply market inefficiency. In this sense, the findings suggest that market efficiency should be considered together with the institutional and behavioral mechanisms through which prices are formed.

The ESG results add a further dimension to this interpretation. Although high-ESG firms exhibit lower clustering in unconditional comparisons, the substantial reduction in the ESG coefficient after controlling for firm and trading characteristics suggests that ESG status does not operate independently of the broader microstructure environment. Differences in firm size, volatility, liquidity, and trading activity appear to account for an important part of the observed ESG/non-ESG gap. Therefore, the results support the view that systematic price patterns may be related not only to investor preferences but also to the characteristics of the firms and trading environments in which those investors operate.

The findings also have practical implications for investors and regulators. For investors, the persistence of clustering around salient price points suggests that round numbers may continue to influence order placement and price formation, even in a decimalized market with a finer price grid. Investors should therefore recognize that observed price patterns may reflect both behavioral tendencies and differences in liquidity, volatility, and firm characteristics. At the same time, ESG classification should be considered alongside these characteristics when assessing differences in price formation across stocks. For regulators, the persistence of clustering is relevant to assessing the quality of price formation and whether market outcomes are shaped by trading frictions, investor behavior, and market structure. The results also suggest that differences across groups of securities should be evaluated after accounting for their underlying characteristics.

The findings should be interpreted considering the limitations of the study. First, ESG classification is time-invariant and is based on the score available in May 2025, meaning that firms are assigned to the same ESG group throughout the 2020-2025 sample period. The analysis therefore does not capture changes in ESG classification over time. Second, the study uses daily closing prices rather than transaction-level or order-book data, which limits

the capacity to identify the specific trading mechanisms that generate clustering. Third, ESG classification is used as an indirect way of examining potential differences in investor preferences, since the dataset does not contain direct measures of investor ownership, order flow, or investor type. Therefore, the analysis cannot establish that the observed differences are caused specifically by investors with ESG-related preferences. In addition, the sample focuses on firms in the highest and lowest ESG-score quartiles with market capitalization above USD 1 billion, so the results may not generalize fully to smaller firms or to companies with intermediate ESG scores. Finally, the empirical design is associational rather than causal, and unobserved firm or market characteristics may still contribute to the relationships identified in the analysis.

Future research could use time-varying ESG classifications, transaction-level data, and direct measures of investor ownership or order flow to examine more directly whether ESG-related investor preferences influence price formation and price clustering.

References

Aitken, M., Brown, P., Buckland, C., Izan, H. Y., & Walter, T. (1996). Price clustering on the Australian Stock Exchange. *Pacific-Basin Finance Journal*, 4(2-3), 297-314. https://doi.org/10.1016/0927-538X(96)00016-9

Aşçıoğlu, A., Comerton-Forde, C., & McInish, T. H. (2007). Price clustering on the Tokyo Stock Exchange. *Financial Review*, 42(2), 289-301. https://doi.org/10.1111/j.1540-6288.2007.00172.x

Baig, A. S., Blau, B. M., & DeLisle, R. J. (2022). Does mutual fund ownership reduce stock price clustering? Evidence from active and index funds. *Review of Quantitative Finance and Accounting*, 58, 615-647. https://doi.org/10.1007/s11156-021-01004-0

Ball, C. A., Torous, W. N., & Tschoegl, A. E. (1985). The degree of price resolution: The case of the gold market. *Journal of Futures Markets*, 5, 29-43. https://doi.org/10.1002/fut.3990050105

Bhattacharya, U., Holden, C. W., & Jacobsen, S. (2012). Penny wise, dollar foolish: Buy-sell imbalances on and around round numbers. *Management Science*, 58(2), 413-431. https://doi.org/10.1287/mnsc.1110.1364

Bollen, N. P. B. (2007). Mutual fund attributes and investor behavior. *Journal of Financial and Quantitative Analysis*, 42(3), 683-708. https://doi.org/10.1017/S0022109000004142

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

Christie, W. G., & Schultz, P. H. (1994). Why Do NASDAQ Market Makers Avoid Odd-Eighth Quotes? *Journal of Finance*, 49(5), 1813-1840. https://doi.org/10.1111/j.1540-6261.1994.tb04782.x

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., García Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27-46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Easley, D., & O'Hara, M. (1987). Price, trade size, and information in securities markets. *Journal of Financial Economics*, 19(11), 69-90. https://doi.org/10.1016/0304-405X(87)90029-8

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. *Management Science*, 60(11), 2835-2857. https://doi.org/10.1287/mnsc.2014.1984

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Goodhart, C. A. E., & Curcio, R. (1991). *The clustering of bid/ask prices and the spread in the foreign exchange market* (Financial Markets Group Discussion Paper No. 110). Financial Markets Group, London School of Economics and Political Science. https://www.fmg.ac.uk/publications/discussion-papers/clustering-bidask-prices-and-spread-foreign-exchange-market

Gormley, T. A., & Matsa, D. A. (2014). Common errors: How to (and not to) control for unobserved heterogeneity. *The Review of Financial Studies*, 27(2), 617-661. https://doi.org/10.1093/rfs/hht047

Hameed, A., & Terry, E. (1998). The effect of tick size on price clustering and trading volume. *Journal of Business Finance & Accounting*, 25(7-8), 849-867. https://doi.org/10.1111/1468-5957.00216

Harris, L. (1991). Stock price clustering and discreteness. *The Review of Financial Studies*, 4(3), 389-415. https://doi.org/10.1093/rfs/4.3.389

Hartzmark, S. M., & Sussman, A. B. (2019). Do investors value sustainability? A natural experiment examining ranking and fund flows. *The Journal of Finance*, 74(6), 2789-2837. https://doi.org/10.1111/jofi.12841

He, Y., & Wu, C. (2006). Is stock price rounded for economic reasons in the Chinese markets? *Global Finance Journal*, 17(1), 119-135. https://doi.org/10.1016/j.gfj.2006.06.008

Hong, H., & Kacperczyk, M. (2009). The price of sin: The effects of social norms on markets. *Journal of Financial Economics*, 93(1), 15-36. https://doi.org/10.1016/j.jfineco.2008.09.001

Hu, B., Jiang, C., McInish, T., & Zhou, H. (2017). Price clustering on the Shanghai Stock Exchange. *Applied Economics*, 49(28), 2766-2778. https://doi.org/10.1080/00036846.2016.1248284

Ikenberry, D. L., & Weston, J. P. (2008). Clustering in U.S. stock prices after decimalisation. *European Financial Management*, 14(1), 30-54. https://doi.org/10.1111/j.1468-036X.2007.00410.x

Kavajecz, K. A., & Odders-White, E. R. (2004). Technical analysis and liquidity provision. *The Review of Financial Studies*, 17(4), 1043-1071. https://doi.org/10.1093/rfs/hhg057

Lin, M., Lucas Jr, H. C., & Shmueli, G. (2013). Research commentary-too big to fail: large samples and the p-value problem. *Information systems research*, 24(4), 906-917. https://doi.org/10.1287/isre.2013.0480

Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. *The Journal of Finance, 72*(4), 1785-1824. https://doi.org/10.1111/jofi.12505

Liu, H.-C., & Witte, M. D. (2013). Price clustering in the U.S. Dollar/Taiwan Dollar swap market. *Financial Review*, 48, 77-96. https://doi.org/10.1111/j.1540-6288.2012.00353.x

Lobão, J., Pacheco, L., & Carvalho, D. (2024). Exploring the Nordic numbers: An analysis of price clustering in Scandinavian stocks. *Review of Behavioral Finance*, 16(6), 1012-1028. https://doi.org/10.1108/RBF-01-2024-0007

McLaren, C. E., Legler, J. M., & Brittenham, G. M. (1994). The generalized χ^2 goodness-of-fit test. *The Statistician*, 43(2), 247-258. https://doi.org/10.2307/2348342

Narayan, P. K., Narayan, S., Popp, S., & D'Rosario, M. (2011). Share price clustering in Mexico. *International Review of Financial Analysis*, 20(2), 113-119. https://doi.org/10.1016/j.irfa.2011.02.003

Niederhoffer, V. (1965). Clustering of Stock Prices. *Operations Research*, 13(2), 258-265. https://doi.org/10.1287/opre.13.2.258

Niederhoffer, V., & Osborne, M. F. M. (1966). Market Making and Reversal on the Stock Exchange. *Journal of the American Statistical Association*, *61*(316), 897-916. https://doi.org/10.1080/01621459.1966.10482183

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, 41, 673-690. https://doi.org/10.1007/s11135-006-9018-6

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22, 435-480. https://doi.org/10.1093/rfs/hhn053

Riedl, A., & Smeets, P. (2017). Why do investors hold socially responsible mutual funds? *The Journal of Finance*, 72(6), 2505-2550. https://doi.org/10.1111/jofi.12547

Uberti, L. J. (2017). Stata Tip 128: Marginal effects in log-transformed models: A trade application. *The Stata Journal*, 17(3), 774-778. https://doi.org/10.1177/1536867X1701700315

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects. *The Stata Journal*, 12(2), 308-331. https://doi.org/10.1177/1536867X1201200209

FEP FACULDADE DE **ECONOMIA E GESTÃO**