The price of simplicity

Ignoring Financial Fundamentals in Housing Markets

PHILIP JACOBSSON

VIKTOR SJÖSTRÖM

Master Thesis

Stockholm School of Economics

2026

The price of simplicity: Ignoring financial fundamentals in housing markets

Abstract

The Swedish cooperative apartment market is one of the largest asset markets where pricing is determined by individuals. More notably, a large share of the apartments constituting this market are organised as cooperative apartments, where the buyer acquires not only the right to use an apartment but also an indirect exposure to the financial position of the tenant-owner association (TOA). This raises the question of whether publicly available TOA financial information is reflected in transaction prices, or whether buyers substitute simpler, more salient indicators for the underlying balance sheet. Using a temporally matched sample of approximately 230,000 Swedish cooperative apartment transactions between 2012 and 2026, linked to TOA annual reports through a novel AI-assisted data pipeline, we analyze the potential effects of TOA financial performance and reporting quality reflected in the pricing of tenant-owned apartments both nationally and across regional submarkets. We find that buyers price the visible monthly fee strongly but do not independently price association debt once the fee is controlled for. Stockholm municipality is a clear exception, where debt and reporting quality enter prices in ways more consistent with balance-sheet capitalisation. Across the rest of Sweden, accounting quality and net debt appears to reach prices only through the monthly fee.

Key words:
Swedish Housing Market, Tenant-Owner Associations, Financial Fundamentals, Accounting Quality, Hedonic Pricing Theory, Limited Attention, Information Processing, Agentic Workflows, AI Extraction, Document Retrieval

Authors:
Philip Jacobsson (42776)
Viktor Sjöström (25472)

Tutor:
Henrik Nilsson, Professor, Department of Accounting

Master thesis
Master Program in Accounting, Valuation and Financial Management
Stockholm School of Economics
Philip Jacobsson and Viktor Sjöström, 2026

Acknowledgements
We would like to express our sincere gratitude to our supervisor, Henrik Nilsson, Professor at the Stockholm School of Economics, for his guidance, support, and valuable feedback throughout this thesis.

Stockholm, May 2026

Philip Jacobsson Viktor Sjöström

Contents

1. Introduction

1.1 Background

The Swedish housing market constitutes a central component of both household wealth and the broader economy. Real estate accounts for approximately two thirds of global net worth, with residential property alone comprising 46 per cent of that total (Woetzel et al., 2021). Despite its size and importance, pricing on the housing market is largely determined by private individuals rather than professional investors, which distinguishes it from many other large asset markets. This raises fundamental questions regarding how financial information is processed, how risks are perceived, and whether housing prices fully reflect underlying economic fundamentals.

In Sweden, a large share of apartments are organised as cooperative apartments, where the buyer acquires not only the right to use an apartment but also an indirect exposure to the financial position of the tenant-owner association (TOA). This exposure includes the TOA's assets and liabilities, implying that an apartment may carry an indirect debt through the TOA's balance sheet. A distinguishing institutional feature of this market is that membership in a Swedish TOA is, in the majority of cases, restricted to private persons. While Swedish law does not impose an absolute prohibition on corporate ownership of cooperative apartments, associations are granted an unconditional right (Bostadsrättslagen, SFS 1991:614) to deny membership to legal entities in dwellings not intended for recreational use. In practice, the overwhelming majority of Swedish TOAs exercise this right (HSB, 2021), meaning that the buyers responsible for pricing cooperative apartments are almost exclusively private individuals. This has direct consequences for how financial information enters market prices, particularly given that housing markets are already characterised by considerably more noise in pricing than markets dominated by professional investors (Case and Shiller, 1988).

In equity markets, professional analysts continuously process financial disclosures and incorporate their implications into prices through competitive trading. Even an unsophisticated investor benefits indirectly from this mechanism, as prices already reflect what analysts have read and assessed. No equivalent mechanism exists in the cooperative apartment market, since no apartments are identical. Each buyer must independently locate, read, and interpret the annual report of the relevant TOA in order to assess its financial health before making what is, for most households, the largest financial decision in their lives. Whether the average apartment buyer is equipped to do this is far from obvious. Research on household financial literacy documents that a large share of individuals struggle with basic financial concepts such as debt sustainability, compound interest, and the evaluation of long-term financial obligations (Lusardi and Mitchell, 2014). Applied to the cooperative housing context, this raises a pointed question if the buyers responsible for setting prices can reliably interpret financial statements.

This question sits at the intersection of accounting research and housing economics, and it has received surprisingly little attention. The accounting literature has extensively studied how financial reporting quality affects pricing efficiency in equity markets, showing that higher quality disclosure reduces information asymmetry and improves the accuracy with which fundamental value is reflected in prices (Diamond and Verrecchia, 1991; Biddle et al., 2009). These findings, however, presuppose that recipients of financial disclosures can process and act on them. The Swedish cooperative housing market is particularly demanding in this respect, as buyers must assess not only the apartment itself but also their indirect exposure to the financial position of the association, encompassing both liabilities and assets on the TOA balance sheet. In a market populated exclusively by private individuals with limited financial training, this assumption may not hold.

A common feature of observed housing market behaviour is the use of simplified indicators when assessing affordability and financial risk. In the context of TOA apartments, the monthly fee is often treated as a key metric by buyers, as it is prominently displayed on listing platforms such as Hemnet and Booli alongside physical attributes such as size and price, making it immediately observable at zero search cost. The financial fundamentals underlying it, namely debt levels, interest rate exposure, operating cost structure, and maintenance reserves, are by contrast embedded in annual reports that must be actively sought out, downloaded, and interpreted. Whether buyers substitute this simple and salient indicator for the more demanding task of evaluating the underlying financial position of the association, is what this paper sets out to examine. However, even among buyers who do engage with TOA annual reports, the extent to which debt is reflected in prices may depend on how clearly and transparently the association presents its financial information, making accounting quality a further dimension of the pricing problem.

This issue has gained renewed relevance in recent years due to changes in the macroeconomic environment. After a prolonged period of near-zero interest rates, Sweden experienced a sharp increase in borrowing costs from 2022 onwards, as the Riksbank raised its policy rate from zero to four per cent over the course of 2022 and 2023 in response to elevated inflation (Sveriges Riksbank, 2026). Higher interest rates directly affect both household mortgages and the debt servicing costs of TOAs, increasing the importance of debt levels and interest rate sensitivity. As financing conditions tightened, assumptions that were relatively inconsequential in a low-rate environment have had material consequences for TOA finances as rates increased sharply.

In Norway, sellers of cooperative apartments are legally required to disclose the buyer's proportional share of collective association debt separately in all property listings (Eiendomsmeglingsloven, 2007), and under the Norwegian lending regulation, banks are required to include joint debt in the calculation of a buyer's total loan-to-value ratio when assessing mortgage affordability (Utlånsforskriften, 2020). This means that debt is made visible to buyers as a matter of course, already at the point of sale. The contrast with Sweden, where debt is embedded in annual reports rather than surfaced in listings, illustrates how institutional design shapes whether financial obligations are visible to buyers at all. A further institutional feature is that Swedish mortgage regulation focuses on the buyer's direct mortgage relative to the apartment's transaction value, rather than on the buyer's indirect exposure to debt through the TOA. During the sample period, the mortgage cap limited direct mortgage borrowing to 85 per cent of the home's market value, while association-level debt was not incorporated into this loan-to-value measure (FI, 2010). This distinction may reinforce the perception that indirect TOA debt is separate from the individual buyer's own leverage.

1.2 Purpose and research questions

The purpose of this paper is to examine the extent to which financial information is reflected in apartment transaction prices in the Swedish cooperative housing market. Specifically, we investigate if buyers incorporate TOA financial fundamentals into transaction prices, and if accounting quality, defined as the clarity and transparency with which a TOA presents its financial information, is directly reflected in apartment transaction prices.

1.3 Delimitations

The analysis is restricted to tenant-owned apartments in Sweden and does not extend to owner-occupied houses or rental properties, which operate under fundamentally different institutional arrangements. The sample covers transactions between 11 March 2012 and 21 March 2026, a period that includes both the low interest rate environment of the 2010s and the sharp rate increases from 2022 onwards (Sveriges Riksbank, 2026). The study examines price formation from the buyer's perspective and does not model seller behaviour or negotiation dynamics.

2. Literature review and theoretical framework

2.1 Hedonic pricing

Hedonic pricing theory provides the foundation for analysing housing markets. Rosen (1974) showed that the price of a differentiated good is determined by the bundle of characteristics it contains. In the housing market, an apartment is not a homogeneous product. Instead, its price reflects a combination of attributes such as size, location, number of rooms, building quality, and neighbourhood characteristics. Each attribute has an implicit price, representing buyers' willingness to pay for a marginal change in that feature (Rosen, 1974).

A key assumption of hedonic pricing theory is that relevant and observable attributes are incorporated into prices in market equilibrium (Rosen, 1974). If buyers are aware of a characteristic and understand its implications, the market price should adjust accordingly. However, this assumption depends on the availability and clarity of information. As discussed by Zabel (2015), capitalisation depends not only on economic relevance but also on whether buyers are able to observe and interpret the information correctly.

In the context of cooperative housing, financial characteristics such as housing association debt can be viewed as attributes affecting ownership costs and financial risk. If higher association debt implies higher expected future fees or greater financial vulnerability, theory predicts a negative relationship between debt and apartment prices (Rosen, 1974).

Empirical evidence broadly supports this reasoning. Narwold et al. (2018) find that cooperative units with higher homeowner association fees sell at lower prices than otherwise comparable units. Their results suggest that buyers negatively capitalise monthly fees into transaction prices. Similarly, research comparing different ownership structures indicates that institutional arrangements influence valuation. Robertsen and Theisen (2011), analysing the Norwegian market, show that cooperative apartments trade at a discount relative to non-cooperative apartments, suggesting that buyers take financial structure and associated risks into account.

2.2 Housing association debt and Swedish evidence

From a valuation-theory perspective, the Modigliani and Miller (1958) irrelevance proposition holds that under perfect capital markets, capital structure does not affect the value of the underlying asset, only how the claims on it are divided. Applied to a tenant-owner association, the starting identity is $V(E) = V(A) - V(L)$, where $V(E)$ is the aggregate residual value accruing to members, $V(A)$ the fair value of total assets, and $V(L)$ the fair value of total liabilities. Following the standard decomposition into operating and financial activities (Feltham and Ohlson, 1995), this can be rewritten as $V(E) = V(ONA) - V(ND)$, where $V(ONA)$ is operating net assets[1] and $V(ND)$ is net debt[2]. For a given $V(ONA)$, every SEK[3] of

[1] Operating assets less operating liabilities, dominated in this context by the property value
[2] Financial debt less cash and cash equivalents
[3] SEK = Swedish krona, the official currency of Sweden

net debt is therefore one SEK less of residual value attached to each apartment. Association debt is serviced through monthly fees, which makes the fee a natural price signal. But the fee is a current cash-flow measure, not a balance-sheet one. It captures only today's interest service at the prevailing rate, scheduled amortisation, and operating expenses. The debt stock is a separate variable. It determines how much the interest cost can rise when fixed-rate periods end, and how thin the equity buffer is when major maintenance is due. Two associations with identical fees are therefore not financially equivalent if one carries substantially more debt. Under rational valuation, buyers should price association net debt directly into apartment values, independently of the fee. Yet empirical findings suggest that net debt effects are often weaker than predicted by full capitalisation, which is itself informative about the frictions in practice.

Evidence from Sweden is limited but illustrates this pattern. Meyer and Ulmgren (2018), in a bachelor's thesis from the Stockholm School of Economics, find that buyers only partially incorporate association debt into prices, with the estimated effect substantially smaller than what full capitalisation would imply. They also report that monthly fees are more strongly reflected in prices than debt itself. This suggests that buyers may respond more readily to visible and easily understood cost measures than to balance-sheet figures that require more interpretation. Related evidence is provided by Hjalmarsson and Hjalmarsson (2006), who study the Swedish cooperative apartment market and find that buyers do not fully capitalise the present value of future monthly fees into transaction prices. Their findings indicate that even recurring ownership costs, which are more visible than balance-sheet debt, may be underweighted in pricing. Taken together, these studies suggest that financial obligations embedded in Swedish cooperative apartments are reflected in prices only partially, and that the degree of capitalisation appears to depend on how salient and interpretable the information is to buyers.

2.3 Accounting quality and information asymmetry

While previous research has examined the role of association debt and monthly fees, less attention has been given to the quality of financial reporting itself. In this context, accounting quality refers to how clearly, accurately, and transparently a housing association presents its financial information. This includes the clarity of annual reports, balance sheets, notes on debt structure, and disclosures regarding future maintenance or financial risks. Associations differ in how detailed and understandable their reporting is, and this variation may influence how buyers assess risk.

From an information asymmetry perspective, differences in reporting quality can have important consequences. Akerlof (1970) shows that when one party in a transaction has better information than the other, uncertainty can lead to price discounts or mispricing. In cooperative housing, financial statements are technically available to all parties, but interpreting them requires a certain level of financial understanding. If financial information is complex or poorly structured, buyers may struggle to evaluate the true implications of association debt. This can create an effective information gap between informed insiders, such as board members or financially sophisticated buyers, and less experienced buyers.

Higher accounting quality can reduce this information gap. Clear and transparent reporting makes it easier to assess debt sustainability, future renovation needs, and financial stability. In financial markets, improved disclosure has been shown to reduce information asymmetry and improve pricing efficiency (Diamond and Verrecchia, 1991). Botosan (1997) provides direct empirical support for this mechanism, demonstrating that firms with higher quality disclosures face a lower cost of capital, as transparent reporting reduces the estimation risk and uncertainty borne by investors. By analogy, clearer reporting in housing associations may reduce the uncertainty buyers face when assessing financial risk, leading to higher willingness to pay and more accurate capitalisation of balance-sheet information into transaction prices.

Evidence from the housing market also supports the importance of information clarity. Berg (2026) shows that when simplified financial ratings were introduced for Swedish housing associations, apartment prices adjusted more strongly around rating thresholds. Before the ratings were introduced, financial conditions had weaker price effects. This suggests that when information becomes easier to interpret, it has a clearer impact on market prices. Accounting quality can be viewed as a similar mechanism operating internally, as associations that provide clearer and more structured financial information may facilitate more accurate pricing.

Bokvist and Lanner (2018), in a Master's thesis from the Stockholm School of Economics, provide related Swedish evidence by examining whether the adoption of the K3 framework improved financial reporting quality among the largest 100 tenant-owner associations and whether TOA accounting information is reflected in apartment prices. They find that voluntary disclosure and reporting harmonisation improved following K3 adoption, but also conclude that apartment buyers do not appear to use unfiltered accounting information to its full extent. Their findings therefore support the view that accounting quality may matter for price formation, while also suggesting that the usefulness of accounting information depends on whether buyers are able to interpret and act on it.

At the same time, low financial literacy among households may increase the importance of reporting quality. Research shows that many individuals struggle to interpret financial concepts related to debt and long-term obligations (Lusardi and Mitchell, 2014). In the context of cooperative housing, this implies that complex or unclear financial reporting may lead some buyers to ignore debt-related risks. In contrast, transparent and well-structured reporting may help bridge this gap and enable more informed decisions.

2.4 Behavioural foundations

A recurring finding in the behavioural economics literature is that individuals do not process all available information when making complex decisions. Instead, attention is a scarce cognitive resource that is selectively allocated to the most accessible and interpretable inputs.

Sims (2003) formalises this intuition through the theory of rational inattention, showing that agents facing information-processing constraints optimally choose to ignore information that is costly to acquire or interpret, even when that information is economically relevant. In settings where decision-makers are not professionals and where the complexity of available information is high, such selective attention is particularly likely to shape outcomes.

Housing purchase decisions represent a natural context for these dynamics. Buyers are typically private individuals rather than trained analysts, and the decision involves evaluating a large and heterogeneous set of attributes. In the cooperative housing setting, the full financial assessment of a TOA requires locating, reading, and interpreting annual reports and what financial consequences it will have for the buyer. This creates a high cognitive cost of processing financial information, which rational inattention theory predicts will lead buyers to underweight or disregard such information in favour of simpler indicators.

Bordalo, Gennaioli and Shleifer (2013) provide a complementary framework through salience theory, which predicts that attributes which are more visible and immediately comparable receive disproportionate weight in decision-making, independent of their objective importance. In the Swedish housing market, the monthly fee occupies precisely this role. It is prominently displayed on listing platforms such as Hemnet alongside physical attributes such as size and price, making it immediately observable without any additional search effort. By contrast, TOA debt levels, interest rate exposure, and maintenance reserves are embedded in annual reports that must be actively sought out, downloaded, and interpreted. Salience theory therefore predicts that the monthly fee will be overweighted relative to underlying financial fundamentals because it is the most cognitively accessible financial signal available. Kahneman (2011) describes this substitution process as a general feature of intuitive judgement, when a question is difficult to answer, people unconsciously replace it with an easier one.

Hirshleifer and Teoh (2003) show that investors selectively attend to information that is salient and easily interpretable, and that the same underlying financial information can have different price effects depending on how prominently it is presented. Importantly, they demonstrate that this tendency to focus on surface-level indicators rather than underlying fundamentals persists even when the financial consequences of ignoring that information increase. Lusardi and Mitchell (2014) document that a large share of individuals lack the basic financial knowledge required to evaluate concepts such as compound interest, debt sustainability, and long-term financial obligations, a pattern that is particularly pronounced among individuals with lower income and education levels (van Rooij, Lusardi and Alessie, 2011). In the cooperative housing context, this implies that even when annual reports are publicly available and of high quality, many buyers may lack the tools to extract meaningful information from them.

2.5 Geographical differences

Behavioural mechanisms are unlikely to operate uniformly across markets. Hedonic pricing theory predicts that implicit prices of housing attributes can vary across submarkets, as buyers in different markets may weight the same attribute differently depending on local conditions (Goodman and Thibodeau, 1998). Financial literacy and buyer sophistication are not uniformly distributed across the population, and income and education levels vary meaningfully across Swedish regions (SCB, 2024), with research consistently documenting that financial knowledge is positively associated with both (Lusardi and Mitchell, 2014; van Rooij, Lusardi and Alessie, 2011). To the extent that financial fundamentals constitute a housing attribute in the hedonic sense, their implicit price may therefore vary systematically across markets with different buyer compositions, suggesting that the degree to which they are capitalised into prices may vary geographically.

2.6 Theoretical framework and hypothesis development

Hedonic pricing theory predicts that economically relevant and observable attributes are incorporated into prices in market equilibrium (Rosen, 1974), and the capital structure irrelevance theory that balance-sheet items carry pricing relevance independent of current cash flows (Modigliani and Miller, 1958). TOA net debt is such a balance-sheet item, an obligation against the cooperative's assets that rational buyers, engaging with the accounting information available to them, ought to price independently of the monthly fee. Because the monthly fee reflects only current debt service costs and not the interest rate risk embedded in the debt stock itself, two associations with identical monthly fees are not financially equivalent if one carries substantially more debt. Under rational valuation, buyers should therefore capitalise the monthly fee negatively into prices (H1), and should additionally price net debt independently and negatively after controlling for the fee (H2).

As discussed earlier, the Swedish cooperative apartment market, however, is populated exclusively by private individuals rather than professionally trained investors. This structural feature has direct implications for how accounting information enters prices. Rational inattention theory suggests that private buyers facing high cognitive costs of processing annual reports will optimally allocate their attention elsewhere (Sims, 2003), while salience theory indicates that the monthly fee, immediately observable on listing platforms at zero search cost, will attract disproportionate weight relative to the balance-sheet information embedded in documents that must be actively sought out and interpreted (Bordalo, Gennaioli and Shleifer, 2013). Limited financial literacy further increases this barrier, as many buyers may lack the tools to extract meaningful information from accounting disclosures even when these are of high quality (Lusardi and Mitchell, 2014). Under these conditions, net debt is likely to be reflected in prices only to the extent that it is captured indirectly through the monthly fee, leaving no independent association with price once the fee is controlled for. Similarly, accounting quality may have limited moderating reach, not because disclosure is irrelevant in principle, but because the private individuals responsible for pricing apartments are not engaging with accounting information in the way that standard theory assumes.

The extent to which cognitive frictions dominate may however depend on the quality of financial reporting available to buyers. Accounting quality can reduce the information asymmetry between associations and buyers, with clearer and more transparent reporting enabling more accurate interpretation of debt levels (Diamond and Verrecchia, 1991; Botosan, 1997). The disclosure quality literature predicts both a direct effect and a moderating effect. The direct effect holds that higher quality reporting reduces estimation risk and information asymmetry, generating a transparency premium as buyers are willing to pay more for associations whose finances are clearly presented (H4). The moderating effect holds that when financial information is clearer and more interpretable, buyers are better able to assess the implications of association net debt, leading to stronger negative capitalisation of net debt into prices (H5).

An additional prediction concerns the macroeconomic environment. The Riksbank's rapid rate increases from near zero to four per cent beginning in May 2022 (Riksbanken, 2026) raised the cost of TOA debt, with debt costs in 2024 approximately three times higher than in 2021 (Nabo, 2025). Salience theory predicts this should have made debt levels more visible and more financially consequential for prospective buyers (Bordalo, Gennaioli and Shleifer, 2013). This generates the prediction that the negative capitalisation of debt into prices should strengthen after May 2022 (H3), though the competing prediction from persistent fee anchoring, that buyers may continue to attend primarily to the monthly fee even as its underlying debt burden becomes more costly, remains theoretically plausible (Hirshleifer and Teoh, 2003).

The empirical predictions follow directly from this framework. If accounting information is processed rationally, net debt should carry an independent negative price effect conditional on the monthly fee, and accounting quality should strengthen debt capitalisation. If cognitive frictions dominate, the monthly fee should absorb the debt signal and accounting quality should show limited moderating effect. As discussed in Section 2.5, these predictions are not expected to be uniform across markets. Income and educational levels differ meaningfully across Swedish regions (SCB, 2024). These differences imply that the degree to which buyers engage with TOA annual reports is unlikely to be uniform across the country. For this reason, all five hypotheses are tested not only on the full pooled sample but also across regional subsamples, to examine whether the degree to which financial information enters apartment prices varies systematically across Swedish housing markets.

2.7 Hypothesis summary

Hypothesis	Definition
H1	Higher monthly TOA fees are negatively capitalised into apartment transaction prices
H2	TOA debt is negatively capitalised into apartment prices beyond what fees already capture
H3	The capitalisation of TOA debt into apartment prices strengthens after the May 2022 Riksbank rate-tightening cycle, when the carrying cost of debt becomes salient to buyers
H4	Higher accounting quality of TOA financial reporting is positively associated with apartment transaction prices
H5	Higher accounting quality of TOA financial reporting shifts the debt-price slope in the direction of balance-sheet capitalisation

Note: All hypothesis are tested across the whole of Sweden, but also across regional subsamples

3. Methodology

The following chapter describes the empirical design, the construction of the matched dataset, and the econometric specifications used in the analysis. We begin by outlining the research setting before turning to the sample construction. We describe the two data sources that together form the analytical sample, namely apartment transaction records obtained from Booli and association-level financial information extracted from tenant-owner association (TOA) annual reports. Because most associations have historically been exempt from public filing requirements, the latter data collection was compiled through an AI-assisted collection and extraction pipeline. Given the non-standard nature of this process and its centrality to the analysis, the retrieval, extraction, and validation procedures are documented in detail.

3.1 Empirical design and research setting

The study adopts a quantitative, observational research design to examine whether variation in TOA financial characteristics and reporting quality is associated with apartment transaction prices after controlling for unit attributes and fine-grained location and time effects. The analysis is conducted within a hedonic pricing framework in which transaction prices are modelled as a function of apartment attributes together with TOA-level financial and accounting-quality variables.

3.2 Data sources and sample construction

The analytical sample is constructed by merging transaction-level data from Booli with association-level financial data compiled from TOA annual reports. The two datasets and the procedure used to link them are described in this chapter. Prior research on hedonic pricing in a Swedish tenant-owner setting has typically faced a trade-off between sample breadth and informational depth. Studies prioritising depth have manually extracted accounting information from annual reports but at a substantially smaller scale, most notably Bokvist and Lanner (2018), who restrict their analysis to the 100 largest TOAs with roughly 20,000 transactions. Studies prioritising breadth rely on third-party data providers and the narrow financial variables these sources provide, such as Hjalmarsson and Hjalmarsson (2006) with over 30,000 transactions from Värderingsdata AB and Meyer and Ulmgren (2018) with around 150,000 Booli transactions. The AI-based workflow used in this study is able to combine both dimensions, producing a Swedish sample with both greater scale and greater financial detail than previous studies, comprising roughly 230,000 regression-ready observations.

3.2.1 Transaction data

Transaction-level data on apartment sales were obtained from Booli, which, to our knowledge, operates the most comprehensive publicly accessible repository of Swedish apartment transaction records. The dataset was retrieved on 11 March 2026 and covers sales concluded between 21 March 2012 and 10 March 2026, resulting in an initial sample of 1,212,483 observations. Each observation corresponds to a single transaction and contains detailed unit attributes, including but not limited to living area, floor level, number of rooms, and the final transaction price.

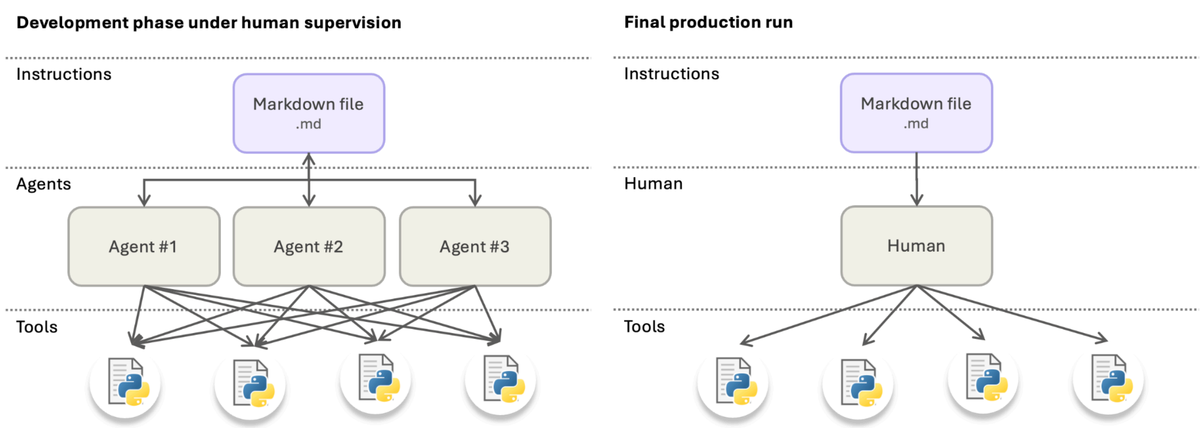
3.2.2 TOA annual report data

Public access to annual reports for Swedish TOAs is severely constrained, and even the institutional databases accessible through the Stockholm School of Economics lacked systematic coverage of these filings. This gap stems from the historical reporting regime administered by the Swedish Companies Registration Office (Bolagsverket), which, pursuant to the Årsredovisningslag (1995:1554), required annual report submissions only from associations classified as large enterprises under the statutory thresholds for employees, turnover, and balance-sheet totals. Because the overwhelming majority of Swedish TOAs are residential entities falling below these thresholds, most were exempt from public disclosure obligations (Bostadsrättsägarna, 2024). Recent legislative reforms extend the filing mandate to all TOAs for fiscal years commencing on or after 1 January 2025, but these filings will not enter the public record until later in 2026 and therefore lie outside our sample window.

These institutional constraints create several practical challenges for data collection. Filings are dispersed across thousands of independent association websites and third-party property-management portals, with no central repository from which to harvest them. Within each site, the target documents are typically buried several levels deep in the navigation hierarchy and are reachable only through idiosyncratic menu sequences that vary across portals. Once located, the reports display substantial heterogeneity in document structure and reporting disclosure. Conventional scraping pipelines, which need a stable page template and predictable document schema, therefore scale poorly in this setting. We accordingly implement an agentic data retrieval system.

Three-layer agentic extraction workflow

Figure 3.1. Three-layer framework for agentic execution



The retrieval system was in the development phase organised into three layers with an instructions layer, an agent layer, and a tool layer, as illustrated in Figure 3.1. This structure is consistent with recent agentic-orchestration literature, where central coordinators delegate subtasks to specialised agents equipped with task-specific instructions, context, and tools (Erik S. & Zhang, 2024; Ruan et al., 2026). The instructions layer consisted of a markdown file specifying the task sequence, branching conditions, tool calls, and stopping criteria. The agent layer, powered by Claude, interpreted these instructions and coordinated the execution of tools. The tool layer consisted of Python scripts that performed deterministic operations including search, crawling, downloading, file handling, and interaction with external services.

Development proceeded iteratively from an initial high-level task specification encoded in the workflow file (Appendix A.1). Multiple agents worked in parallel to generate tool implementations and progressively refine the workflow design, enabling rapid exploration at a breadth that would have been significantly more time-consuming to achieve through manual development alone. These autonomous runs served as the primary mechanism for identifying bugs, improving fetching and extraction logic, and revealing weaknesses in the workflow specification. The generated tools and resulting outputs were continuously reviewed through human intervention. This human-in-the-loop refinement cycle is characteristic of semi-automated data collection in research contexts (Artemova et al., 2025).

Although the framework allowed autonomous end-to-end execution, the final production run was executed manually rather than through the agents. Agents were valuable during development for tool creation, workflow refinement, debugging, and end-to-end testing. However, sustained autonomous operation over a multi-day retrieval pipeline would have required repeated model calls and long context histories, which recent work identifies as a major cost driver in multi-step agent systems (Xiao et al., 2026). Manual execution also improved quality assurance, since retrieval-based pipelines are vulnerable to compounding errors (Xie & Sun, 2026). Once validated, manual execution therefore offered a more cost-efficient and controllable production strategy.

Figure 3.2. Overview of data retrieval process for TOA annual report data

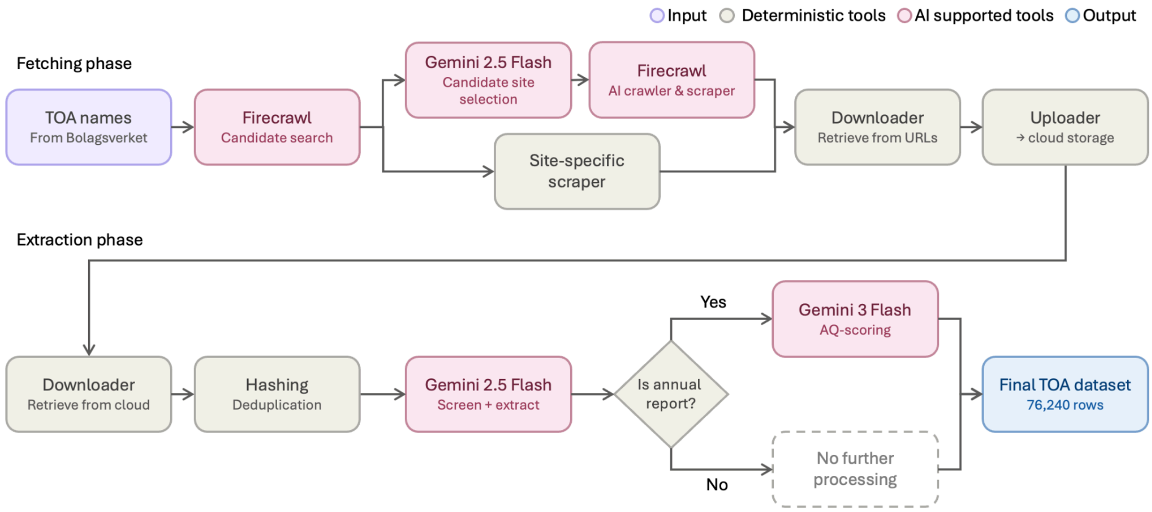


Figure 3.2 provides a high-level overview of the final retrieval workflow, whose main steps are discussed in the following subsections.

Using TOA names as input to the workflow

Names of active TOAs are obtained from Bolagsverket and supplied as queries to the later search stage. The decision to query by association name rather than by organisational number emerged through trial and error during early retrieval development, with name-based queries consistently producing more useful results. Since search engines do not rely on exact string matches but retrieve pages on the basis of approximate name similarity, the approach absorbs the spelling variations that commonly arise between the canonical name registered with Bolagsverket and the form used on sites hosting association documents. A further advantage was that name-based queries more often returned candidate URLs pointing directly to association webpages or other publicly accessible sources from which annual reports could be downloaded, whereas queries built on organisational numbers tended to surface generic registry entries and external contact information pages of little use for retrieval.

Fetching phase - Search and retrieval procedure

Search and crawling were mainly implemented through Firecrawl, a web data API for AI-oriented search and crawling. It was used to run web queries, return candidate URLs, and crawl selected pages for annual-report links. Its principal advantage was providing search and crawling through a unified interface, reducing the custom scraping logic that would otherwise have had to be developed and maintained across heterogeneous websites. It also supported JavaScript-rendered navigation and multi-level internal link discovery, which proved important as annual reports were often located several layers below landing pages and labelled inconsistently across sites.

For each query, Firecrawl returned ten candidate sites, each with a URL, title, search-result position, and short site description. Direct PDF links and standardised websites were separated and handled through dedicated retrieval paths, as discussed below. Certain sites, such as *hitta.se* and *Eniro*, were excluded as they frequently appeared as candidates but contained no reports. The remaining candidates were evaluated by Gemini 2.5 Flash to identify the most likely official website for each TOA (see Appendix A.2). This AI-based ranking was used because URL structures, domain names, and page labels varied substantially across associations and property-management portals. Unlike a rule-based approach, the model could weigh contextual signals such as association-name similarity, description, link labels, and organisational relevance, making selection more flexible than deterministic ranking. The selected page was sent back to Firecrawl for further crawling, and URLs related to annual reports were saved to a separate file documenting which to visit during the later downloading stage. The purpose of this filtering and selection layer was to avoid spending resources on implausible, duplicate, or low-value retrieval paths. Clearly irrelevant results and weak candidates were filtered out before deeper crawling, which was important given the project's time and budget constraints, since deep crawling requires substantially more processing and token usage. Although crawling every candidate in depth could in principle have improved recall, this was not feasible within the available constraints.

The retrieval procedure therefore prioritised plausibility-guided selection over exhaustive coverage.

For a small subset of standardised websites, site-specific tools were introduced instead of general agentic crawling. These tools used traditional scraping logic tailored to each website's structure. This reduced cost by avoiding unnecessary model-based searches, improved accuracy by targeting the relevant document links directly, and increased speed as execution could be run locally. Such adaptations were only implemented where the site structure was sufficiently regular to justify a tailored approach. Found URLs were posted to the output file.

Direct PDF links returned by Firecrawl were written directly to the output file. Together with URLs identified through crawling and site-specific tools, these links formed the standardised set of candidate annual-report URLs passed to the download and storage stage. As this URL set was populated, a dedicated download tool retrieved the linked PDF files and uploaded them to the cloud, reducing the risk of data loss from local interruptions during the multi-day collection process and to accommodate the large volume of reports flowing in.

Extraction phase - Hashing, validating and extracting

Once reports were retrieved, the pipeline entered the extraction phase, consisting of hashing, screening and financial extraction, and accounting-quality scoring. Each PDF was assigned a content-based digital fingerprint via hashing. Hashing helped to identify duplicate files to prevent them from being sent to later stages, but also allowed interrupted runs to resume without reprocessing previously handled files. It was also used to point to the source file.

Google's Gemini 2.5 Flash-Lite was used as a combined screener and extractor in a single pass over each document. The model first determined whether the document was a relevant TOA annual report and then extracted the structured financial fields defined in Appendix A.3. Combining both tasks in one call reduced cost and latency, since a two-step design would have required uploading and processing the same PDF twice, approximately doubling model cost and processing time for this stage. Documents classified as non-annual reports were still returned in the predefined JSON schema with empty fields and were discarded in post-processing based on the classification flag.

We constrained the model in several ways. First, the model was required to return JSON rather than free prose, a method commonly applied in this setting (Dagdelen et al., 2024). Second, the model was provided with a fixed extraction template specifying each target field and its expected data type, in line with prior work on structured-format generation for information extraction (Li et al., 2024). Third, the temperature parameter was set to zero to reduce stochastic variation across repeated runs, consistent with established practice in structured LLM[4]-based extraction workflows (Dagdelen et al., 2024). This overall design is consistent with recent methodological work showing that LLMs can extract structured

[4]LLM = Large language model, an AI model designed to process and generate natural language text.

records from unstructured documents with high accuracy when tasks are clearly specified and outputs are constrained, and with evidence that, under predefined codebooks and expert-adjudicated benchmarks, LLM-based extraction can achieve accuracy comparable to human expert extraction (Jansen et al., 2026).

Extraction phase - Accounting Quality (AQ) Scoring

Accounting quality scoring was performed in a separate stage using Google's Gemini 3 Flash, which was selected as the task required evaluative judgement over narrative disclosures rather than only field identification and extraction. Keeping the stages separate allowed extraction errors and rating errors to be examined independently in subsequent validity checks and allowed each prompt to be tailored to its specific task. This separation is consistent with task-decomposition research showing that complex LLM workflows can benefit from being divided into narrower subtasks (Liu et al., 2025). It also reduced cost by applying the more capable scoring model only to documents that passed the initial annual-report screening.

Table 3.1. AQ scoring criteria

Scoring dimensions	Score 0	Score 1	Score 2
Future maintenance disclosure (A)	No information or vague information	Maintenance actions are mentioned, but without timing or cost	Clear maintenance actions are described with timing, cost, financing, or expected fee impact
Risk and financing disclosure (B)	No risks are disclosed, or only generic risks are mentioned	Risks are mentioned, but without clear financial effect or planned response	Clear TOA-specific risks are linked to financial consequences, exposure, or mitigating actions
Important events disclosure (C)	No events are disclosed, or only standard wording is used	Events are mentioned, but without clear consequences	Events are linked to operating result, financial position, liquidity, debt, or future outlook

Note: Text used in the actual prompting is translated from Swedish.

The scoring model applied a predefined rubric constructed as a structured disclosure index, in which each TOA annual report is evaluated against a fixed set of predefined criteria. Disclosure indices are widely used in accounting research to translate narrative disclosures into comparable quantitative measures by evaluating reports against explicit information categories with consistent coding rules (Marston and Shrives, 1991; Beattie, McInnes and Fearnley, 2004). The approach has been particularly prominent in sustainability reporting research, where disclosure indices have proven effective at capturing the qualitative richness of narrative information that simpler quantitative proxies cannot adequately reflect (Gutierrez-Bustamante & Espinosa-Leal, 2022). However, applications of disclosure-based accounting quality measures to Swedish tenant-owner associations remain limited to our knowledge. A relevant exception is Bokvist and Lanner (2018), who convert TOA annual-report disclosures into quantitative measures of financial reporting quality. Our approach builds on similar logic, but focuses more directly on buyer-relevant interpretability and information rather than formal reporting harmonisation around K3 adoption. A central premise of the methodology is that informational value depends on specificity and decision relevance rather than disclosure volume (Beretta and Bozzolan, 2004), which motivates the scoring design, where a score of 2 requires concrete, quantified, time-bound, or economically linked disclosure rather than a mere mention of the topic. See Table 3.1 for scoring criteria.

The three dimensions, future maintenance disclosures (A), risk and financing disclosures (B), and material events and their economic consequences (C), are grounded in Swedish accounting regulation for TOA financial reporting. Under BFNAR 2023:1[5], TOAs are required to disclose information about future maintenance and reinvestment needs, including timing and estimated cost, how future financial obligations such as interest costs and debt service will be financed, and material circumstances affecting the association together with their economic consequences for the association's financial position (Bokföringsnämnden, 2023). Given that limited research exists on disclosure idenx in the cooperative housing context specifically, grounding the dimension selection in the regulatory framework established by the authoritative standard-setter provides a principled and externally validated basis for identifying which categories of information are most decision-relevant for assessing TOA accounting quality. Associations scoring highly on the index approach the informational standard the regulatory framework requires, while associations scoring poorly fall short of what the standard-setter identifies as necessary for informed decision-making.

For each dimension, the model assigned an ordinal score from 0 to 2 and reported an uncertainty score. Full scoring criteria and score distributions are reported in Appendix A.4. The output was returned in a predefined JSON schema, and the model was instructed to assign the lower score whenever the report did not provide sufficient textual support for the higher rating. Following the same constraint logic applied at the extraction stage, this design is consistent with recent LLM-evaluation evidence emphasising that model-based scoring is more reliable when rubrics are fixed, criteria are explicit, and outputs are constrained (Hong et al., 2026; Saez et al., 2026).

Extraction and scoring validation

To assess extraction quality, we manually verified a randomly selected sample of 30 processed annual reports against the corresponding source PDFs. The AI extracted fields were checked field-by-field against the values reported in the original documents. No discrepancies were identified in the reviewed sample. This result does not establish that the full dataset is error-free, but it provides evidence from a limited manual validation check in which no extraction errors were detected. The same procedure was applied to the scoring stage, with no evidence of misclassification in the reviewed sample. Given the high skewness in the accounting-quality scores, three additional reports were randomly drawn from reports

[5] BFNAR 2023:1 refers to Swedish Accounting Standards Board's (*Bokföringsnämnden*) general guidance on supplementary disclosures in Swedish housing cooperative annual reports.

scoring 0 and three from reports scoring 1. The reviewed annual reports are consistent with the observed skewness in accounting-quality scores, although the limited validation sample should be interpreted as a plausibility check rather than independent evidence of the full-sample distribution.

3.2.3 Matching procedure and analytic sample

We link each transaction in Booli to the tenant-owner association that owns the underlying property. We then assign to the transaction the most recent annual report from that association that would have been publicly available at the time of sale. The association-level link relies on the organisational number as the join key, whereas the transaction-to-report assignment is governed by a temporal availability rule described below. The organisational number is a stable statutory identifier assigned by Bolagsverket and provides an unambiguous reference to a specific legal entity. Association names, by contrast, are recorded inconsistently across sources where the same entity may appear as, for example, *BRF Flugan* in one source and *Bostadsrättsföreningen Flugan i Leksand* in another. This renders name-based matching unreliable at scale when information is retrieved from different sources.

The Booli dataset does not contain organisational numbers. Each transaction record does, however, link to a Booli-internal webpage for the corresponding TOA, and these pages, for a substantial share of associations, host the association's annual reports. We therefore recover the identifier from the reports themselves rather than from the Booli metadata. Swedish annual reports are statutorily required to disclose the association's organisational number under Chapter 2, Section 5 of the Annual Accounts Act (1995:1554), and this information is, by convention, placed on the cover page, in the management report, or in other early sections of the document. Because the location of the disclosure varies across associations while its presence is mandated, the first pages of each report are a reliable extraction target, whereas the layout heterogeneity within those pages motivates the use of an AI-based extraction step rather than a rule-based parser if done at scale. Specifically, we apply Google's Gemini 3.0 Flash Lite model to the first three pages of each retrieved report to locate and extract the organisational number. Restricting the input to the first three pages reflects a trade-off between cost and performance. On the cost side, processing full reports to recover a single identifier is substantially more expensive than processing a short prefix, and the marginal gain from additional pages is small given that the identifier is, in the overwhelming majority of cases, disclosed on the opening pages. On the performance side, a narrower input reduces the scope for spurious matches and is consistent with evidence that smaller context windows improve accuracy on narrow identification tasks (Du et al., 2025).

The choice of Gemini 2.5 Flash Lite reflects the narrow scope of the task. Because the objective is to locate a single, well-defined identifier rather than to interpret complex structures requiring deep reasoning, a lighter model is deemed sufficient. This choice is consistent with results from ExtractBench, where Google's lighter models perform competitively on constrained extraction tasks and, in several cases, outperform larger frontier models (Ferguson et al., 2026).

To assess extraction accuracy, we randomly sample 50 reports and manually verify the extracted identifier against the source document. The model's output matches the document in all 50 cases. After full extraction, minor formatting inconsistencies are observed, as organisational numbers appear in varying representations with respect to hyphens and spaces. We normalise all entries to the standard six-plus-four digit format used by Bolagsverket through a deterministic Python post-processing routine. The collected reports were also processed through the extraction phase of the pipeline, as described in Section 3.2.2.

Temporal matching procedure

We restrict matches to annual reports that would plausibly have been available to market participants at the time of sale. Establishing when a report is available requires a date that applies uniformly across associations, which the filing regime alone cannot supply. As noted in Section 3.2.2, most TOAs fell below the statutory size thresholds and were therefore not required to file their annual reports with Bolagsverket during our sample period. The Economic Associations Act (2018:672), which governs TOAs by cross-reference from the Housing Cooperative Act (1991:614), supplies an approximation to such a date through two requirements. Chapter 6 §9 requires every association to hold its ordinary general meeting within six months of the fiscal year end, and the annual report must be presented and adopted at that meeting. Chapter 6 §23 requires the signed report to be made available to members no later than two weeks beforehand. For any associations that additionally exceed the thresholds of the Annual Accounts Act (1995:1554) and are consequently obliged to file, the filing deadline is seven months post-fiscal-year-end. This is one month beyond the general meeting deadline. Taking the later of the two deadlines, we treat the availability date of each annual report as the fiscal year end plus seven months, which is the earliest date by which the report is statutorily guaranteed to be available for every association in our sample, regardless of filing requirements. While one could argue that an earlier release is practically possible, the general meeting date is unobservable in our dataset and we therefore apply this cautious approach. This conservative timing assumption is also practically reasonable because the annual report must be prepared, reviewed, signed, made available to members, and formally adopted before or at the ordinary general meeting, making immediate availability after the fiscal year-end unlikely.

Each transaction is matched to the most recent annual report from the same association whose implied availability date precedes the transaction date. The relevance window of each report is clipped at the implied publication date of its successor, so that once a newer report becomes publicly available the older one is no longer eligible to match. This procedure gives at most one report per transaction and enforces temporal consistency between the financial information used in the specification and the information plausibly available to the buyer. For example, a transaction occurring in March 2025 would be matched to the 2023 annual report as the 2024 report had not yet become publicly available according to our matching principle. By contrast, a transaction occurring in August 2025 would be matched to the 2024 annual report as the report is now assumed to be available.

However, a remaining concern is that the prior-year report is occasionally assigned where the current-year report has in fact become available, whether through the one-month slack between the general meeting and filing deadlines or through variation across associations in how adopted reports are circulated to members and listing platforms. The relevant question is the magnitude of the induced error rather than its elimination. Aggregate association-level debt per square metre for Swedish TOAs moves only marginally year-on-year, with fixed amortisation averaging SEK[6] 47 per square metre against an average debt stock of SEK 6,221 per square metre, implying an annual reduction on the order of 0.8 per cent of the outstanding balance in the typical association (Fredrik Fexe & Berglind, 2026). Naturally, individual associations deviate from this average through renovation borrowing, refinancing, or front-loaded repayment, but such cases are a minority and the residual measurement error is therefore bounded. The prior-year report can therefore be treated as a close approximation of the financial information available to buyers at the time of transaction, and the residual noise is unlikely to materially affect the estimated relationships.

Matched dataset characteristics

Of the 1,212,483 transactions in the Booli dataset, 855,758 (70.6 per cent) were linked to a TOA organisational number. Applying the temporal matching procedure reduced the matched sample to 291,007 observations (see Table 3.2). The resulting sample spans 21,706 associations and 56,835 reports, with a median of 3 transactions tied to each report. After further investigation, one should note that the cumulative retention dropped partly due to inconsistent temporal matching across the dataset (see Appendix A.5).

Table 3.2. First stage sample construction

Stage	Filter	Transactions retained	Cumulative retention
0	Raw Booli transactions, 2012-2026	1,212,483	100.0%
1	Linked to TOA organisational number	855,758	70.6%
2	Relevant fiscal-year report available	291,007	24.0%

3.2.4 Location assignment

The choice of location fixed effects involves a trade-off between absorbing meaningful within-area heterogeneity and retaining enough observations within each cell for precise estimation. In the hedonic framework, prices should be estimated within well-defined local housing markets as overly broad geographies risk combining distinct submarkets and biasing the coefficients of interest (Rosen, 1974). However, overly narrow geographies can leave too little variation for reliable estimation. Motivated by evidence that residential prices vary substantially at sub-municipal scales, we use DeSO 2025[7] as the primary spatial fixed effect because it captures fine local variation while preserving usable cell sizes. Postal-code fixed

[6]SEK = Swedish krona, the official currency of Sweden
[7] DeSO = DeSO (Demografiska statistikområden) are small geographic areas defined by Statistics Sweden (SCB) for regional statistics.

effects are used as a robustness specification at a comparable level of granularity, and Table 3.3 reports how each candidate geography performs against these two constraints.

Table 3.3. Comparison of location fixed-effect specifications

Level	Unique cells	Median obs / cell	Cells < 30 obs (%)	Within R^2
Municipality	256	80	35.9%	0.595
Postcode	680	14	60.4%	0.628
DeSO	3,838	27	51.5%	0.732
Street address	70,740	2	99.7%	0.733

Note: Within R^2 from Eq. (4) (see section 3.4.4), varying location fixed effect only. Street address fixed effects are defined as unique city-by-street-address cells to account for the fact that a street can have the same name in multiple cities.

Municipality fixed effects may be too coarse for the purposes of this application. In Stockholm municipality alone, mean price per square metre at the postal-code level ranges from 36,529 SEK at the 10th percentile to 119,039 SEK at the 90th percentile, indicating substantial within-municipality price heterogeneity. Consistent with this, the municipality fixed-effects specification yields an in-sample R^2 of 0.595 after controlling for structural property characteristics and time effects, suggesting that meaningful variation remains unaccounted for. Street-address fixed effects, however, may be too granular for the purposes of this paper. The sample contains 70,740 unique street cells, with a median of only two transactions per cell, and 99.7 per cent of cells contain fewer than 30 observations. The 30-observation cutoff is used as an illustrative benchmark rather than a formal econometric requirement. Samples of around 30 observations are commonly treated as a rule-of-thumb threshold at which normal approximations based on the central limit theorem become more reliable, although this threshold is context-dependent and should not be interpreted mechanically (Kwak and Kim, 2017). In this setting, the issue is not that OLS requires 30 observations per group, but that sparsely populated street cells provide limited within-street variation for estimating fixed effects. As a result, street-address fixed effects estimated from very few transactions may absorb idiosyncratic transaction-level variation in addition to persistent local price differences, making them less suitable as the main spatial fixed-effect specification, particularly when the fixed-effect structure creates singleton or near-singleton groups (Correia, 2016).

DeSO 2025 units occupy the middle ground on both counts. Introduced by Statistics Sweden in 2018 as demographically stable small-area geographies of 700-2,700 residents, designed to nest within municipal boundaries, DeSO partitions are explicitly engineered for the demographic homogeneity for statistical purposes. Across 3,838 DeSO cells in our sample, the median cell contains 27 transactions nationally. Replacing municipality fixed effects with DeSO fixed effects raises the in-sample R^2 from 0.595 to 0.732, an increase of 13.7 percentage points. Expressed relative to the residual variance under the municipality specification, this represents a reduction of approximately 33.8 per cent, suggesting that DeSO fixed effects capture a substantial amount of within-municipality spatial price variation. Although DeSO is specific to Sweden, using similarly fine-grained location fixed effects, such as postcodes, is common in hedonic housing research (Heyman and Sommervoll, 2019). We therefore also include postal codes in our sample for robustness purposes.

Neither DeSO codes nor postal codes are present natively in the transaction data. DeSO codes are matched using each transaction's geographic coordinates to geographical boundaries provided by SCB. Similarly, we recover postal codes by matching each transaction's geographic coordinates to the nearest Swedish postal-code centroid. This is done via a k-dimensional tree using a reference file of Swedish postal-code centroids obtained from GeoNames. The procedure does not recover the administratively correct postal code for every property near a boundary, but it maps each observation to a well-defined local neighbourhood.

3.2.5 Data cleaning and correction

To correct obvious anomalies in the matched dataset prior to estimation we conducted a final data-cleaning procedure. Because a significant number of variables in the dataset are parsed by AI models from heterogeneous PDF reports, some degree of field-level miscoding, unit inconsistency, and outlier generation is expected at this scale. We therefore conduct a manual review targeting values that are economically implausible, internally inconsistent across statements, or demonstrably inconsistent with the underlying annual report. Corrections were made only where the source of the error could be identified manually.

The largest correction category concerns financial-statement scaling errors. In a subset of association-year observations, annual-report values had been parsed at the wrong monetary scale, most commonly because amounts reported in thousands of Swedish kronor were extracted as if they were Swedish kronor, or because SEK-denominated values were treated as TSEK-denominated values. For each affected association-year, we rescaled the relevant financial-statement variables after verification against the source report. The correction was restricted to balance-sheet items with non-monetary variables from the report left untouched. In total, the script defines 130 organisation-year unit-scale corrections, of which 101 involve multiplying the targeted balance-sheet items by 1,000 and 29 involve dividing the same set of variables by 1,000.

The second category consists of source-verified point corrections to isolated erroneous values. These include 24 corrections to apartment living area, 11 corrections to total debt, and one correction to the number of rooms. We identified these cases through a manual review of values that were suspicious, internally inconsistent, or economically implausible. Examples include observations in which apartment living area was recorded as zero, apartment-level and association-level area measures appeared to have been conflated, or an apartment characteristic had evidently been entered in the wrong field. Finally, we applied three rule-based cleaning adjustments after the manual corrections where negative values for total debt were converted to absolute values, apartment living areas equal to zero or above 500 square metres were set to missing, and monthly fees recorded as zero were likewise set to missing.

The subsequent estimation sample is subject to additional pre-specified validity restrictions and limited percentile trimming. We first impose logical bounds that exclude observations inconsistent with a valid transaction or association-level financial statement, including non-positive sale prices, non-positive apartment living area, negative monthly fees, and non-positive association area. Following Zabel (2015), we apply a symmetric 1% trim to Fee and Debt (section 3.4), the two ratios most directly built from AI-extracted financial-statement inputs and thus most vulnerable to errors. The same reasoning motivates trimming the non-owner area ratio (section 3.3.2), though only at the upper tail. This variable is bounded below at zero, and low values are economically meaningful, whereas implausibly high values likely reflect extraction or denominator errors. Trimming is preferred to winsorising in this setting because the concern is not that valid extreme values should be compressed towards the centre of the distribution, but that a small number of observations may reflect residual construction or scaling errors. Removing only the most extreme tails of these two constructed ratios is therefore a transparent way to limit the influence of potentially erroneous observations while preserving the remaining observed variation. This trim is not applied to all variables, nor is it intended to remove economically inconvenient observations. Finally, all main regressions are estimated on the same common sample, defined by the most restrictive specification, Eq. (7). This means that only transactions with complete valid data for all variables used in the full model are retained. Using a common sample ensures that differences across specifications reflect changes in model structure.

Since extreme values in housing data may reflect genuine variation rather than measurement error, we trim only the three AI-derived financial ratios exposed to residual extraction or scaling errors. This selected trimming intends to preserve economically meaningful cross-sectional variation and limits mechanically generated outliers, consistent with guidance cautioning against excessive outlier removal (Osborne and Overbay, 2004; Aguinis et al., 2013). The final regression ready sample contains 229,940 transactions, across 17,703 unique associations and 45,976 reports. See Table 3.4 for second stage sample construction.

Table 3.4. Second stage sample construction

Stage	Filter	Transactions retained	Cumulative retention
0	Base sample (Table 3.2)	291,007	100.0%
1	Apartments only	283,092	97.3%
2	After logical bounds	265,716	91.3%
3	After 1.0% trim of *Fee*, *Debt*, and *NA ratio*[8]	252,649	86.8%
4	After common sample	229,940	79.0%

Note: Second stage sample selection builds from table 3.2.

[8]*Non-owner area ratio* is total rental and premises area divided by area with tenant owner right, capturing association space not allocated to owner-occupied apartments (see section 3.3.2).

3.3 Variables

3.3.1 Dependent variable

The dependent variable in all main specifications is the logarithm of the apartment sale price i.e. *ln(Price)*. The logarithmic transformation is common in hedonic pricing models and is applied because housing prices are typically right-skewed and because many housing characteristics are expected to affect prices proportionally rather than by a constant SEK amount (Zabel, 2015). We use total transaction price rather than price per square metre because the latter imposes a unit elasticity between price and apartment area. Apartment size is instead included as a separate control, allowing the price-size relationship to be estimated directly.

3.3.2 Control variables

Although the dataset contains a large number of potential variables, the empirical specifications prioritise controls with high data availability, high theoretical relevance and empirical support in previous Swedish and international hedonic housing-price research (Hjalmarsson & Hjalmarsson, 2006; Meyer and Ulmgren, 2018; Zabel, 2015). The carefully selected variables and their justification are provided next. Further statistics on control variable availability is reported in Appendix A.6.

Log living area

Following Zabel (2015), apartment size is controlled for using *ln(living area)*. Including it helps ensure that estimated effects from other variables are not simply capturing differences in apartment scale. Since both transaction price and living area are log-transformed, the coefficient on living area can be interpreted as the percentage change in price associated with a one per cent increase in living area, holding other variables constant.

Number of rooms

The variable *rooms* controls for the number of rooms in the apartment. Conditional on living area, the number of rooms captures layout intensity. For example, two apartments with the same area may differ in market value depending on whether the space is divided into more or fewer rooms.

Floor

The variable *floor* controls for the floor level of the apartment. Floor level may affect price through light, noise exposure, views, accessibility, and perceived attractiveness.

Qualified housing association (Äkta bostadsrättsförening[9])

The variable *Qualified housing association* is a binary indicator coded one when the annual report identifies the TOA as *äkta* (äkta bostadsrättsförening) or as a *privatbostadsföretag*, and zero when identified as *oäkta*. The classification is included as a control because it affects both the tax treatment of apartment ownership and the association's underlying income structure. Non-qualified associations are subject to different rules for taxation upon sale and potential annual taxation of housing benefits (Skatteverket, n.d.), which can affect the net economic value of ownership and be reflected in transaction prices. The classification also relates to the share of revenue derived from commercial premises and other non-member uses, which may influence monthly fees and perceived financial stability (Fastighetsägarna, 2017; Bostadsrätterna, 2022). Controlling for this therefore helps isolate the variables of interest from differences in the legal, tax, and financial conditions linked to the association's classification.

Leasehold (Tomträtt)

Leasehold is a binary indicator equal to one when the TOA annual report indicates that the property is held under a site-leasehold[10] arrangement and zero otherwise. This control is included because leasehold associations pay an annual ground rent to the municipality or public landowner (SBC, 2023b; Stockholms Stad, 2026). Since ground rent represents an ongoing cost and may be renegotiated after a fixed period, leasehold status can affect association fees, expected future costs, and perceived financial risk (Stockholms Stad, 2026; Solna Stad, 2026). Controlling for leasehold status helps separate the effects of the main financial variables from differences in land-tenure arrangements.

Construction-period indicators

The regressions include construction-bin dummy variables based on the construction year of the apartment to account for systematic differences in building age, construction standards, architectural period, renovation needs, and neighbourhood sorting that may be correlated with both TOA finances and prices. Construction-period indicators follow Booli's established construction-year periodisation for Swedish housing. The cutoffs correspond to historically interpretable building eras rather than data-driven breaks. Properties with missing construction year are assigned to a separate unknown category to avoid dropping observations with incomplete construction-year data. See Appendix A.7 for classification details.

[9] Swedish term: "*äkta bostadsrättsförening*", denoting a tenant-owner association that qualifies as a private housing company, primarily relevant for tax purposes.
[10] Site leasehold refers to the Swedish "tomträtt", where the association owns the right to use the land, but not the land itself, and pays ground rent (tomträttsavgäld).

Newly built indicator

New building is a binary indicator equal to one if the building was less than five years old at the time of sale, and zero otherwise. The five-year cutoff draws inspiration from hedonic housing research showing that dwelling age and newness can affect transaction prices (Wilhelmsson, 2008). The variable is included as a pragmatic control for potential newness premia that may not be fully absorbed by the broader construction-period controls.

Non-owner area ratio (NA-ratio)

Non-owner area ratio is defined as rental-apartment area plus premises area divided by area with tenant-owner right[11]. The variable captures association-controlled space not allocated to TOA apartments. From a valuation perspective, an apartment's transaction price should reflect not only the characteristics of the apartment itself and the association's net debt, but also the operating assets the association holds beyond the tenant-owner stock. Rental units and premises generate recurring income, and they can also be converted and sold as TOA apartments, with proceeds available to reduce association debt. Both channels imply a positive contribution to V(ONA) (Feltham and Ohlson, 1995), so that, holding apartment characteristics and net debt constant, a rational buyer should be willing to pay more for membership in an association with greater non-owner operating capacity.

3.3.3 Fixed effects

Location fixed effects

All regressions include location fixed effects to absorb spatially correlated unobservables such as unobserved neighbourhood quality and local amenity provision that would otherwise confound price estimates. The baseline specification employs DeSO 2025 fixed effects (see section 3.2.4). For robustness purposes we also let postal-code fixed effects replace DeSO indicators.

Temporal fixed effects (Year-Month)

To absorb common market-wide shocks, including changes in interest rates, credit conditions, seasonality, macroeconomic expectations, and aggregate housing-market trends, we use temporal fixed effects. This variable is based on the apartment sale date and combines the year and month. In the common regression sample, there are 165 unique temporal cells, ranging from 2012-06 to 2026-03.

[11] Swedish annual report term: “area upplåten med bostadsrätt”, meaning floor area assigned to individual members and not available for the association to rent out or sell. The phenomenon can be compared to shares that are either owned by external shareholders or held by the company itself. Area granted with tenant-owner rights is like shares owned by external shareholders, meaning the other shareholders have no claim to it, while areas that are not granted with tenant-owner rights are like repurchased shares held by the company, meaning shares or space that the remaining shareholders still have a claim to.

3.4 Model specifications

This section describes the empirical models used to test the five hypotheses. The specifications are estimated in a sequential structure, moving from the monthly fee channel to the balance-sheet debt channel, and finally to accounting quality. All regressions are estimated on the same common sample to make the coefficients comparable across specifications.

Inspired by Zabel (2015), the standard hedonic model takes the following form:

$$ln(P_{i,j,t}) = \beta' X_{i,j,t} + \gamma' Z_{j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (1)$$

where $P_{i,j,t}$ is the transaction price of apartment i in association j at time t. $X_{i,j,t}$ contains apartment-level controls, including logarithmic living area, number of rooms, floor, qualifying cooperative status, site leasehold, construction-period indicators and whether the building is newer than five years at the time of sale. $Z_{j,t}$ contains TOA-level financial and accounting variables. $FE_{location}$ denotes DeSO 2025 location fixed effects in the main models, and FE_{time} denotes the year-month fixed effects. Standard errors are clustered at the organisation level to allow for within-organisation correlation in the error terms (Cameron & Miller, 2015).

3.4.1 H1: Fee capitalisation

The first specification tests whether higher monthly TOA fees are negatively capitalised into apartment prices (H1):

$$ln(P_{i,j,t}) = \beta_1 Fee_{i,j,t} + \beta' X_{i,j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (2)$$

where $Fee_{i,j,t}$ is the monthly association fee per square metre of living area for apartment i in TOA j at time t. The coefficient of interest is β_1, and H1 predicts $\beta_1 < 0$. The specification follows the standard hedonic pricing framework, in which apartment prices reflect the bundle of housing characteristics and recurring ownership costs (Rosen, 1974). Conditional on dwelling characteristics, location, and the timing of sale, β_1 captures the association between the monthly fee per square metre and log transaction price.

3.4.2 H2: Balance-sheet capitalisation beyond the fee

The second hypothesis (H2) asks whether TOA net debt is negatively capitalised into apartment prices beyond what the monthly fee already reflects. Prior work suggests that associations with higher debt levels tend to incur higher interest and amortisation costs, which may be passed through to members as higher monthly fees (see discussion in section 2.8). If such a pass-through exists, a reduced-form association between debt and price may capture the fee channel rather than direct balance-sheet pricing. To test whether net debt carries pricing information beyond the fee, two related specifications are estimated on the same common sample, following the approach in the value relevance literature of testing whether an accounting variable is priced beyond other observable signals (Barth, Beaver, and Landsman, 2001). The first is a reduced-form specification that omits the monthly fee:

$$ln(P_{i,j,t}) = \beta_1 Debt_{j,t} + \beta' X_{i,j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (3)$$

where $Debt_{j,t}$ is the net debt per square metre of cooperative apartment area (*bostadsrättsyta*) reported in the annual accounts of TOA j most recently available prior to the transaction at time t. Net debt is calculated as the sum of interest-bearing liabilities less cash and cash equivalents. Debt or net debt in this paper will refer to this definition if not stated otherwise. Eq. (3) is not a separate hypothesis test and is therefore referred to as H2 (red.) in some instances indicating that this is the reduced-form specification. It recovers the unconditional association between net debt and apartment prices, which includes both the indirect channel operating through fees and any direct balance-sheet channel. The second, main H2 specification, adds the monthly fee:

$$ln(P_{i,j,t}) = \beta_1 Fee_{i,j,t} + \beta_2 Debt_{j,t} + \beta' X_{i,j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (4)$$

where the coefficient of interest in equation (4) is β_2, the association between net debt and price conditional on the monthly fee. H2 predicts $\beta_2 < 0$.

3.4.3 H3: Debt capitalisation post May 2022

The third hypothesis (H3) tests whether the association between TOA debt and apartment prices changed following the start of the Riksbank rate-tightening cycle in May 2022. The choice of May 2022 as the break date reflects the Riksbank's first policy rate increase following a prolonged period in which the rate remained at or below zero, as illustrated in Appendix A.8. To test for a change in the debt-price slope around this event, net debt is interacted with a post-2022 indicator. Interacting a binary indicator with a continuous explanatory variable allows the slope on that variable to differ across periods, making the specification suitable for testing whether the debt-price relationship changed after the rate-tightening event (Wooldridge, 2013):

$$ln(P_{i,j,t}) = \beta_1 Debt_{j,t} + \beta_2 Fee_{i,j,t} + \beta_3 (Debt_{j,t} \times Post2022_t) + \beta' X_{i,j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (5)$$

where $Post2022_t$ is an indicator equal to one for transactions on or after 4 May 2022 and zero otherwise. The coefficient of interest is β_3, which captures the change in the debt-price slope following the rate shock. The standalone $Post2022_t$ indicator is excluded as it is fully absorbed by the year-month fixed effects. H3 predicts $\beta_3 < 0$, suggesting the capitalisation of TOA debt into apartment prices strengthens after the May 2022 Riksbank rate-tightening cycle, when the carrying cost of debt becomes salient to buyers.

3.4.4 H4: Transparency premium from accounting quality

The fourth hypothesis (H4) tests whether accounting quality carries a standalone association with apartment transaction prices, conditional on fee and debt.

$$ln(P_{i,j,t}) = \beta_1 Fee_{i,j,t} + \beta_2 Debt_{j,t} + \beta_3 AQ_{j,t} + \beta' X_{i,j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (6)$$

$AQ_{j,t}$ is the composite accounting quality score for TOA j at time t, ranging from 0 to 6. As indicated in Appendix A.4, the individual variability of the three disclosure-dimension scores may be limited if they are entered separately. We therefore aggregate them into a composite AQ measure as a precaution, increasing the observable variation in the regressor and reducing the risk that estimates are driven by sparse variation. Fee and debt are retained to ensure that β_3 captures any association between accounting quality and price that is not already reflected in the cash-flow and balance-sheet channels. The coefficient of interest is β_3. H4 predicts $\beta_3 > 0$, consistent with the prediction that higher quality accounting information reduces estimation risk and information asymmetry, leading to higher asset prices (Botosan, 1997).

3.4.5 H5: Accounting quality as moderator of debt capitalisation

The fifth hypothesis (H5) tests whether accounting quality moderates the extent to which TOA net debt is capitalised into apartment prices in the direction of balance-sheet capitalisation. H5 examines whether accounting quality changes the pricing weight placed on balance-sheet debt. This follows the disclosure-quality literature, which predicts that accounting information becomes more value relevant when it is more reliable, transparent and easier for market participants to process (Bushman and Smith, 2001; Francis, LaFond, Olsson and Schipper, 2005).

To test this mechanism, net debt is interacted with the composite accounting quality score:

$$ln(P_{i,j,t}) = \beta_1 Fee_{i,j,t} + \beta_2 Debt_{j,t} + \beta_3 AQ_{j,t} + \beta_4 (Debt_{j,t} \times AQ_{j,t}) + \beta' X_{i,j,t} + FE_{location} + FE_{time} + \varepsilon_{i,j,t} \quad (7)$$

where $Debt_{j,t} \times AQ_{j,t}$ is the interaction between TOA net debt per square metre and accounting quality. The coefficient of interest is β_4. Since debt is interacted with accounting quality, β_2 captures the debt-price slope when $AQ_{j,t} = 0$, while the marginal association between debt and price at a given level of accounting quality is:

$$\partial\, ln(P_{i,j,t}) / \partial\, Debt_{j,t} = \beta_2 + \beta_4 AQ_{j,t} \quad (8)$$

H5 predicts $\beta_4 < 0$. This implies that higher accounting quality shifts the debt-price slope in the direction of stronger balance-sheet capitalisation.

3.4.6 Geographic sample decomposition

In addition to the pooled specifications, we estimate each hypothesis at two geographic levels to investigate heterogeneity across Sweden's largest apartment markets. This is motivated by the theoretical expectation that buyers' attention to TOA financial information may vary geographically due to differences in information-processing costs, financial literacy, and economic incentives to examine annual reports (Sims, 2003; Lusardi and Mitchell, 2014; SCB, 2024).

The first level is the county. We re-estimate the main specifications separately for Stockholm County, Västra Götaland County, and Skåne County, which together account for 75.6 per cent of the final regression-ready sample and contain Sweden's three largest metropolitan housing markets. If signs and magnitudes remain similar across the three subsamples, the pooled findings are less likely to be an artefact of Stockholm's large sample weight (47.6%). If the estimates differ across the three subsamples, the pooled coefficient should be interpreted as a variance-weighted average that masks heterogeneous local relationships rather than a uniform national association.

The second level is within-region. Within each of the three counties, we separate the core municipality from the rest of the county (see Table 3.5). This split follows the same spatial logic as the county decomposition but focuses on within-region differences between dense urban cores and surrounding local markets.

Table 3.5. Regional decomposition within the largest three county subsamples

Region	Observations	TOAs	Share within region	Share of full sample
Stockholm County	109,344	6,466	100%	47.6%
Stockholm municipality	44,831	3,784	41.0%	19.5%
Rest of county	64,513	2,682	59.0%	28.1%
Västra Götaland	32,195	2,535	100%	14.0%
Gothenburg municipality	23,042	1,353	71.6%	10.0%
Rest of county	9,153	1,182	28.4%	4.0%
Skåne County	32,280	2,498	100%	14.0%
Malmö municipality	19,876	980	61.6%	8.6%
Rest of county	12,404	1,518	38.4%	5.4%

Note: Table displays regional distribution within the final regression-ready sample across and within the three largest regions Stockholm County, Västra Götaland and Skåne County. Shares denote the % of transactions.

3.4.7 Robustness checks

Several robustness checks were conducted to assess the stability of the specifications. These include alternative spatial fixed effects, alternative debt measures, placebo break dates, Covid-period exclusions, influential-observation checks, within-building quality check, alternative trimming thresholds, and accounting-quality component decompositions. These checks are reported in more detail in Appendix A.9.

Moreover, variance inflation factors (VIFs) are reported for the main specifications to assess the extent to which the estimated coefficients may be affected by problematic multicollinearity among the regressors (Kutner et al., 2018). These diagnostics are complemented by the Pearson correlation matrix, which reports the raw pairwise correlations among the main continuous regression variables (Schober et al., 2018).

4. Result / data

4.1 Descriptive statistics

Table 4.1 reports descriptive statistics for the final estimation sample of 229,940 apartment transactions. The average sold price is approximately 2.84 million SEK, with a median of 2.35 million SEK, indicating a right-skewed price distribution. The median apartment has 2 rooms and is located on the 2nd floor. Monthly association fees average 63.29 SEK per square metre, while average net housing association debt equals 5,403 SEK per square metre. The accounting quality variables are concentrated near their upper values, with the aggregate score averaging 5.73 out of 6. See Appendix A.10 for a Pearson correlation matrix.

Table 4.1. Descriptive statistics

Variable	Obs.	Mean	Median	Std. Dev	Q1	Q3
ln(Price)	229,940	14.66	14.67	0.66	14.30	15.05
Price	229,940	2,839,008	2,350,000	2,044,387	1,625,000	3,450,000
Fee	229,940	63.29	62.33	12.25	54.67	71.10
Debt	229,940	5,403	4,369	4,078	2,181	7,884
AQ	229,940	5.73	6.00	0.88	6.00	6.00
NA ratio	229,940	0.06	0.02	0.08	0	0.08
ln(Living area)	229,940	4.12	4.16	0.36	3.91	4.38
Rooms	229,940	2.49	2.00	0.94	2.00	3.00
Floor	229,940	2.75	2.00	2.08	1.00	3.00
Construction year	229,940	1969	1965	31	1948	1991

Variable definitions: ln(Price) = natural logarithm of apartment transaction price; Price = apartment transaction price in SEK; Fee = monthly association fee per square meter of living area; Debt = housing association net debt per square meter; AQ = aggregate accounting quality score; NA ratio (Non-owner area ratio) = total rental and premises area divided by area with tenant owner right; ln(Living area) = natural logarithm of apartment living area; Rooms = number of rooms; Floor = apartment floor level; Construction year = building construction year.

4.2 Main results

Table 4.2 displays the full sample results, reflecting the countrywide results. As noted in the method section of this paper, the specifications move sequentially from the monthly fee channel to the balance-sheet debt channel, then to post-2022 debt salience, accounting quality, and finally the interaction between accounting quality and debt. All models are estimated on the same common sample and include the full set of controls, DeSO fixed effects, year-month fixed effects, and organisation-level clustered standard errors.

Table 4.2. Hedonic capitalisation of TOA financials

Panel A: Main regression results

Variable	**H1 (Eq.2)**	**H2 (red.) (Eq.3)**	**H2 (Eq. 4)**	**H3 (Eq. 5)**	**H4 (Eq. 6)**	**H5 (Eq. 7)**
Fee	−0.0035***		−0.0037***	−0.0038***	−0.0037***	−0.0037***
Debt		−1.4e−06***	+2.2e−06***	−2.1e−08	+2.2e−06***	+5.8e−06***
Debt × Post2022				+3.8e−06***		
AQ					−0.0005	+0.0036***
Debt × AQ						−6.4e−07***
Controls	Yes	Yes	Yes	Yes	Yes	Yes
DeSO FE	Yes	Yes	Yes	Yes	Yes	Yes
Year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
N	229,940	229,940	229,940	229,940	229,940	229,940
TOAs	17,703	17,703	17,703	17,703	17,703	17,703
Within R^2	0.732	0.724	0.732	0.732	0.732	0.732

*Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.*

Panel B: Implied net-debt slopes

Specification	**Implied slope**
Eq. (5), Pre-2022	−2.1e−08, p = 0.96
Eq. (5), Post-2022	+3.8e−<06***
Eq. (7), AQ = 0	+5.8e−06***
Eq. (7), AQ = 3	+3.9e−06***
Eq. (7), AQ = 6	+2.0e−06***

*Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.*

In hedonic-pricing literature, a within R^2 of approximately 0.73 is not unexpected once apartment-level controls are included and DeSO and year-month fixed effects are absorbed. The within R^2 is also virtually flat across specifications (0.72-0.73), with the largest change in explanatory power occurring when Fee is excluded. Together, these results indicate no substantial risk of overfit in any of the regressions.

4.2.1 Fee capitalisation (H1)

Hypothesis H1 predicts that higher monthly TOA fees are negatively capitalised into apartment transaction prices. Shown in specification 1 of Table 4.2, the coefficient on *Fee* is −0.0035 and statistically significant at the 1% level, supporting that a one-SEK per square metre increase in the monthly association fee is associated with an approximately 0.35% lower transaction price, conditional on controls and fixed effects. This indicates that buyers price recurring ownership costs into transaction values, in line with standard hedonic capitalisation theory.

4.2.2 The wedge between reduced-form and direct debt effects (H2)

Hypothesis H2 predicts that TOA net debt is negatively capitalised into apartment transaction prices beyond what monthly fees already capture. Eq. (3) of Table 4.2 estimates the reduced-form association between net debt and log price without a fee control. The coefficient on *Debt* is -1.4×10^{-6} and statistically significant at the 1% level, indicating that a 1,000 SEK per square metre increase in net debt is associated with an approximately 0.14% lower transaction price in the absence of a fee regressor, conditional on controls and fixed effects. When the monthly fee is introduced in Eq. (4), the coefficient on *Debt* reverses sign to $+2.2 \times 10^{-6}$ and remains statistically significant at the 1% level, indicating that the same 1,000 SEK per square metre increase is associated with an approximately 0.22% higher transaction price conditional on fee. The direct association between net debt and price, holding fee constant, is positive rather than negative meaning no support for H2 is found.

4.2.3 Post-2022 salience (H3)

Hypothesis H3 predicts that the association between TOA debt and apartment prices strengthens following the Riksbank's May 2022 rate-tightening cycle, as the increased cost of debt service raises the salience of TOA indebtedness for prospective buyers. Eq. (5) of Table 4.2 introduces an interaction between net debt and a post-2022 indicator. The interaction coefficient is $+3.8 \times 10^{-6}$ and statistically significant at the 1% level. Panel B of Table 4.2 reports the implied net-debt slopes for each sub-period. Before May 2022, the implied slope is -2.1×10^{-8} and is not statistically significant ($p = 0.96$), indicating no reliably identified association between net debt and price in the pre-shock period. After May 2022, the implied slope is $+3.8 \times 10^{-6}$ and statistically significant at the 1% level, corresponding to an approximately 0.38% higher transaction price per 1,000 SEK per square metre of net debt. H3's prediction of a more negative post-shock debt-price slope is not supported. The evidence instead indicates that the positive direct association between debt and price strengthened following the rate shock.

4.2.4 Accounting quality main effect (H4)

Hypothesis H4 predicts that higher accounting quality of TOA financial reporting is positively associated with apartment transaction prices, reflecting a reduction in information asymmetry and estimation risk. In Eq. (6) of Table 4.2, the coefficient on *AQ* is −0.0005 and is not statistically significant ($p = 0.55$). The evidence does not indicate a positive standalone association between accounting quality and transaction price in the pooled sample. Hence, H4 is not supported.

4.2.5 Accounting quality moderation of debt pricing (H5)

Hypothesis H5 predicts that the association between TOA debt and apartment prices becomes more negative as accounting quality increases, indicating that higher disclosure quality enables greater incorporation of balance-sheet information into transaction prices. In Eq. (7)

of Table 4.2, the coefficient on Debt × AQ is -6.4×10^{-7} and statistically significant at the 1% level, supporting that a one-unit increase in the accounting quality score is associated with a more negative debt-price slope, conditional on controls and fixed effects. Panel B of Table 4.2 reports the implied net-debt slopes at three points of the AQ distribution. At the minimum disclosure score (AQ = 0), the implied slope is $+5.8 \times 10^{-6}$ ($p < 0.001$), indicating an approximately 0.58% higher transaction price per 1,000 SEK per square metre of net debt. At the midpoint (AQ = 3), the slope is $+3.9 \times 10^{-6}$ ($p < 0.001$), corresponding to approximately 0.39%. At the maximum (AQ = 6), the slope is $+2.0 \times 10^{-6}$ ($p < 0.001$), corresponding to approximately 0.20%. As seen in the results, the implied slope declines monotonically across the AQ distribution but remains positive and statistically significant at every observed point. The results therefore indicate that higher accounting quality is associated with a debt-price slope that moves closer to the predicted direction of balance-sheet capitalisation. This supports the directional prediction of H5. However, the implied slopes suggest partial correction rather than full capitalisation. The debt-price association does not turn negative at any observed level of accounting quality, indicating that even high-disclosure TOA transactions do not appear fully consistent with balance-sheet-based debt capitalisation.

4.3 Geographic and urban heterogeneity

The pooled estimates reported in Table 4.2 mask heterogeneity in how TOA financials are priced across Swedish housing markets. As discussed in the method section, this part documents this heterogeneity along two dimensions. First, through the three major counties of Stockholm, Västra Götaland, and Skåne, and second, within each county contrasting between the central municipality and the rest of the region. The geographic patterns identified here inform later interpretation of the conditions under which buyers appear to incorporate balance-sheet information into transaction prices.

Table 4.3. Regional decomposition of headline coefficients

Hypothesis	**H1**	**H2 (red.)**	**H2**	**H3**	**H4**	**H5**
Region	**Fee (Eq. 2)**	**Debt (Eq. 3)**	**Debt (Eq. 4)**	**Debt × Post2022 (Eq. 5)**	**AQ (Eq. 6)**	**Debt × AQ (Eq. 7)**
Full sample	−0.0035***	−1.4e−06***	+2.2e−06***	+3.8e−06***	−0.0005	−6.4e−07***
Stockholm County	−0.0030***	−2.3e−06***	+4.7e−07	+1.8e−06***	−0.0005	−1.1e−06***
Stockholm municipality	−0.0018***	−2.5e−06***	−9.0e−07**	−2.1e−06***	+0.0021*	−7.9e−07***
Rest of county	−0.0040***	−2.5e−06***	+1.1e−06	+5.4e−06***	−0.0011	−7.1e−07***
Västra Götaland	−0.0033***	−8.4e−08	+3.9e−06***	+5.6e−06***	−0.0003	+3.3e−07
Gothenburg municipality	−0.0030***	−2.1e−06***	+1.1e−06	+5.7e−06***	−0.0022	−2.5e−07
Rest of county	−0.0041***	+2.9e−06	+9.2e−06***	+5.4e−06***	+0.0046	+1.1e−06
Skåne County	−0.0039***	−3.7e−06***	+6.3e−07	+2.6e−06***	+0.0006	−5.3e−07
Malmö municipality	−0.0038***	−4.7e−06***	−4.4e−07	+2.5e−06***	+0.0003	−1.4e−07
Rest of county	−0.0046***	−4.3e−06***	+1.0e−06	+1.8e−06	+0.0031	−7.0e−07

*Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.*

The negative association between monthly fees and transaction prices documented in the H1 specification is visible across the three major counties. County-level coefficients range from -3.0×10^{-3} in Stockholm to -3.9×10^{-3} in Skåne and are statistically significant at the 1% level in every region, consistent with H1's prediction of $\beta_1 < 0$. A similar pattern appears at the municipality level, where the fee coefficient is less negative in the urban core than in the surrounding county. It is -1.8×10^{-3} in Stockholm municipality compared with -4.0×10^{-3} in the rest of Stockholm County. Whether this point-estimate pattern reflects genuine differences in fee capitalisation cannot be assessed without formal cross-subsample tests.

The debt pricing pattern in Eq. (3) and Eq. (4) differs across the three major counties. The reduced-form coefficient on net debt in Eq. (3) is negative and significant at the 1% level in Stockholm (-2.3×10^{-6}) and Skåne (-3.7×10^{-6}) but not in Västra Götaland. When fee is held constant in Eq. (4), residual debt is positively associated with price in Västra Götaland ($+3.9 \times 10^{-6}$, $p < 0.01$), directionally inconsistent with H2's prediction of $\beta_2 < 0$, and not statistically distinguishable from zero in Stockholm or Skåne, providing no support for H2 in either county. At the urban level, Stockholm municipality is the only submarket where the direct debt coefficient is both negative and significant (-9.0×10^{-7}, $p < 0.05$), consistent with H2. Gothenburg and Malmö show coefficients close to zero and not statistically significant. The largest divergence occurs in the rest of Västra Götaland, where the direct coefficient is $+9.2 \times 10^{-6}$ ($p < 0.01$), roughly four times the pooled estimate and an order of magnitude larger than the corresponding Gothenburg coefficient, inconsistent with H2. The reduced-form coefficient in that submarket is $+2.9 \times 10^{-6}$ and not statistically significant, indicating that the fee and debt bundle carries no reliable negative association even before fee is controlled for. Within the decomposition, peripheral Västra Götaland returns the largest positive direct debt coefficient ($+9.2 \times 10^{-6}$, $p < 0.01$), though at 4.0% of the full sample it represents a small share of the overall data (see Table 3.5), and this result should be interpreted as a local submarket pattern rather than an indication of broad influence on the pooled estimate.

The interaction between net debt and the Post2022 indicator shows an urban-periphery contrast not visible at the county level alone. The interaction is positive and significant at the 1% level in all three counties ($+1.8 \times 10^{-6}$ in Stockholm, $+5.6 \times 10^{-6}$ in Västra Götaland, and $+2.6 \times 10^{-6}$ in Skåne), inconsistent with H3's prediction of $\beta_3 < 0$. Decomposing into urban cores and peripheries, however, reveals that Stockholm municipality alone carries a negative and significant interaction (-2.1×10^{-6}, $p < 0.01$), indicating that the debt-price slope became more negative after May 2022 in central Stockholm, consistent with H3. The rest of Stockholm County moves in the opposite direction ($+5.4 \times 10^{-6}$), as do the rest of Västra Götaland ($+5.4 \times 10^{-6}$), Gothenburg ($+5.7 \times 10^{-6}$), and Malmö ($+2.5 \times 10^{-6}$), each at the 1% level. The rest of Skåne carries $+1.8 \times 10^{-6}$ and is not statistically significant. Central Stockholm is therefore the only submarket where the post-2022 interaction supports H3, while every other submarket with a reliably identified interaction points in the opposite direction.

The AQ coefficient in the H4 specification is not statistically distinguishable from zero in any county or municipality except Stockholm municipality, where it enters with a weakly significant positive coefficient of $+2.1 \times 10^{-3}$ ($p < 0.10$). The sign matches H4's prediction of $\beta_3 > 0$, providing marginal support for a standalone transparency premium in the central urban submarket. Elsewhere the AQ coefficient remains indistinguishable from zero. The moderation of the debt-price slope by accounting quality appears to be concentrated in Stockholm. The county-level interaction coefficient is -1.1×10^{-6} ($p < 0.01$) in Stockholm, consistent with H5's prediction of $\beta_4 < 0$, but $+3.3 \times 10^{-7}$ in Västra Götaland and -5.3×10^{-7} in Skåne, with the latter two not statistically distinguishable from zero. Within Stockholm County, both the municipality and the rest of the county show significant negative debt-AQ interactions, with similar magnitudes of -7.9×10^{-7} and -7.1×10^{-7}, respectively, each significant at the 1% level. The interaction is therefore supported within the Stockholm region but not in the other two major counties.

Taken together, the geographic decomposition suggests that Stockholm municipality is the submarket whose pricing behaviour is most consistent with the balance-sheet and transparency channels developed in the hypothesis framework. There, the data indicate a negative and significant residual debt coefficient once fee is held constant (Eq. 4), consistent with H2; a negative and significant post-2022 interaction (Eq. 5) showing the debt-price slope became more negative after May 2022, consistent with H3; a weakly significant positive AQ coefficient (Eq. 6) in the direction predicted by H4, providing marginal support for a standalone transparency premium; and a significant negative debt-AQ interaction (Eq. 7) consistent with H5, suggesting accounting quality moderates the debt-price slope towards stronger negative capitalisation. Stockholm municipality is therefore the only submarket where all four balance-sheet and transparency-related hypotheses receive directional support, though evidence for H4 remains marginal at the 10% level. The remaining submarkets show patterns largely inconsistent with H2 and H3, with no support for AQ moderation outside the Stockholm region.

4.4 Robustness

The robustness checks broadly support the main findings, with some qualifications. The fee-capitalisation result appears robust to using postcode rather than DeSO fixed effects, and the debt results remain stable across alternative trim thresholds, gross-debt measurement, and exclusion of high-leverage observations. The post-2022 interaction is supported by the placebo break tests at May 2020 and May 2021, which yield much smaller and insignificant coefficients, and the estimate becomes somewhat larger when the Covid period is excluded. The weak direct AQ effect in H4 persists in the component-wise decomposition, while the H5 moderation result is further supported by the component-specific AQ interaction tests. Within the narrower 2010-or-later subsample, where building quality should be more homogenous, the debt coefficient is small, negative, and insignificant, neither indicating a negative pricing of debt nor reproducing the positive baseline association. This supports that buyers do not systematically capitalise TOA debt into prices, even when building quality is held more tightly constant at the pooled level.

Multicollinearity diagnostic results are reported in Appendix A.10. The VIFs and pairwise correlations suggest no evidence of problematic multicollinearity in the main regressions.

5. Analysis

5.1 Fee anchoring as the dominant pricing mechanism

The significant and consistent negative association between monthly fees and transaction prices documented in H1 suggests that buyers negatively capitalise recurring ownership costs into apartment prices, a finding that holds across all three major counties and is stable across robustness specifications (Eq. 2). This result aligns with standard hedonic capitalisation theory and the prediction of Rosen (1974) that economically relevant and observable attributes are incorporated into market prices.

When TOA net debt is entered into the model without controlling for the monthly fee, the coefficient is negative and statistically significant (Eq. 3). However, once the monthly fee is introduced as a control, the coefficient on net debt reverses sign and becomes positive and statistically significant at the national level (Eq. 4). This is inconsistent with the rational valuation prediction of the capital structure irrelevance theory (Modigliani and Miller, 1958), and instead suggests that buyers treat the monthly fee as a sufficient statistic for total ownership cost, failing to price either the reduced residual net asset value or the embedded interest rate risk that higher debt implies but that the current fee does not capture.

The results indicate that buyers do not engage with TOA annual reports, consistent with rational inattention theory (Sims, 2003). Instead, the monthly fee appears to dominate the valuation process, consistent with salience theory (Bordalo, Gennaioli and Shleifer, 2013), and the H2 results suggest that buyers substitute this single salient indicator for a more complete assessment of the association's financial position, as predicted by Kahneman (2011). In addition, the institutional separation between direct mortgage debt and indirect TOA debt as discussed in the background, may dampen the negative pricing of association debt. Because the Swedish mortgage cap applies to the apartment's transaction price rather than to the buyer's total exposure including proportional TOA debt, high-debt associations may allow financially constrained buyers to access apartments with more attractive physical characteristics at a lower formal purchase price. Association debt can therefore function as indirect leverage outside the formal loan-to-value constraint. This may weaken the negative price effect of net debt, as the structure effectively increases the pool of buyers who can afford apartments in high-debt associations, supporting demand and partially offsetting the negative valuation effect.

However, the positive sign on net debt conditional on fee requires further explanation, as it might superficially appear to suggest that buyers actively reward high-debt associations. We do not interpret the positive coefficient as buyers actively rewarding debt. Several mechanisms outside our data could generate this pattern, of which unobserved renovation quality is the most plausible candidate we believe. Although the regressions control for building age through construction-period indicators, these capture systematic differences across building eras rather than recent capital investments within a given building. We argue this operates through two related mechanisms. First, associations that have recently undertaken major capital investments, such as pipe replacements, roof renovations, or façade upgrades, will typically carry higher debt or lower cash balance as a direct consequence of financing those works. Buyers are likely to value a recently renovated building and may bid higher accordingly, but may not connect that renovation to the increased debt, or decreased cash, that financed it. In doing so, they capture the benefit of the physical improvement without discounting the balance-sheet liability it created, further indicating limited attention to financial statements. Second, and more simply, buyers may rationally pay a premium for the convenience of purchasing a recently renovated property, knowing they are unlikely to face major renovation assessments or fee increases in the near future. Under either mechanism, the positive coefficient does not reflect buyers rewarding debt itself, but rather buyers responding to a visible physical attribute that happens to be correlated with higher debt in the data. Because our dataset does not contain direct measures of recent renovation activity, this quality dimension may not be fully captured in the regression, and its effect may load onto the debt variable, producing an upward bias in the conditional debt coefficient.

However, the suggestion that net debt would turn negative and significant once renovation quality is accounted for is not supported by the new-build robustness test (Appendix A.9). When the sample is restricted to buildings constructed from 2010 onwards, where major structural renovations are unlikely to have occurred (SBC, 2023a), the debt coefficient turns negative but insignificant. The positive sign reversal observed in the full sample therefore appears to be driven by older buildings where renovation history is relevant and unobserved, rather than by buyers placing positive value on debt itself. However, because the 2010-or-later subsample is a restricted and compositionally different part of the market, the result should be interpreted as suggestive rather than conclusive evidence on the renovation-quality channel. Taken together, the H1 and H2 results provide no evidence consistent with rational balance-sheet capitalisation of TOA net debt. The positive conditional coefficient is statistically reliable but, as argued above, is unlikely to represent buyers pricing debt itself, and more plausibly reflects unobserved variation correlated with debt. The fee appears to be the primary vehicle through which debt information reaches prices, and private buyers show no detectable tendency to look beyond it to the underlying balance sheet.

This aggregate pattern, however, masks meaningful regional variation. As documented in Section 4.3, Stockholm municipality is the only submarket in the sample where the direct association between net debt and transaction price is negative and statistically significant once the monthly fee is controlled for, suggesting that some degree of direct debt pricing does occur. This contrasts to the rest of the observable regions where debt seems to operate solely though the fee channel, with the exception of peripheral Västra Götaland, where the conditional coefficient is positive, significant, and roughly four times the pooled estimate. Given the submarket's small share of the data (around 4% of the sample), we treat this as a local pattern rather than evidence of a general mechanism, and do not generalise from it.

The aggregate null result should therefore be interpreted as an average across markets with heterogeneous buyer compositions rather than as a uniform finding across Sweden. This is consistent with the broader financial literacy literature, which documents that a large share of private individuals lack the basic financial knowledge required to evaluate debt obligations, and that this barrier may be particularly consequential in markets where no professional intermediary processes disclosures on behalf of buyers (Lusardi and Mitchell, 2014).

5.2 The post-2022 rate shock and the fee channel

Hypothesis H3 predicted that the capitalisation of TOA debt into apartment prices would strengthen following the Riksbank's rate-tightening cycle beginning in May 2022, as rising borrowing costs increased the financial consequences of association debt for prospective buyers. The results do not support this prediction. The significant interaction coefficient in Eq. (5) captures the positive change in the debt-price slope after May 2022 relative to before. Before the rate shock, the implied debt-price slope is negative but not statistically significant from zero, suggesting no reliably identified association between net debt and transaction prices in the low-rate environment. After May 2022, the implied slope shifts to positive and becomes statistically significant. The direction of this shift is the opposite of what H3 predicted. Rather than debt becoming more negatively priced as its financial consequences grew, the association between debt and price became more positive after the rate shock.

The shift towards a more positive debt coefficient following the rate shock is primarily explained by fee anchoring under delayed cost adjustment. TOAs often hold loans with fixed-rate periods, meaning interest costs adjust gradually as market rates change. Combined with the delay before annual reports become publicly available, the monthly fee at the time of sale may not fully reflect the forward-looking cost implications of the association's debt. A buyer anchoring on the current fee may therefore perceive a high-debt association as no more expensive than a low-debt one with the same fee, without recognising that the former carries greater exposure to future cost increases. That the coefficient turns positive rather than simply remaining near zero is a separate question, and is more likely explained by unobserved renovation quality as discussed in section 5.1.

The post-2022 period is when the financial consequences of association debt were greatest, yet buyers appear to have relied even more heavily on the monthly fee rather than scrutinising the balance sheet more carefully. This is consistent with Hirshleifer and Teoh (2003), who show that the tendency to focus on salient surface-level indicators rather than underlying fundamentals persists even as the financial consequences of doing so increase. In the context of rising interest rates, this mechanism may have been reinforced by stronger affordability constraints. As private mortgage costs increased and household budgets became more constrained, buyers may have focused more strongly on immediate and observable housing costs, particularly their own borrowing costs and the monthly fee. Rather than prompting buyers to engage more directly with TOA financial statements, the rate shock therefore appears to have reinforced fee anchoring, without improving the accuracy of the information the fee conveys.

This interpretation finds support in the geographic decomposition. The post-2022 interaction is positive and statistically significant across all three major counties, suggesting that the rate shock did not prompt more direct balance-sheet engagement at the regional level. However, the county-level regression conceals two exceptions. The rest of Skåne County was positive but not statistically significant. Stockholm municipality carries a negative and statistically significant interaction, indicating that the debt-price slope became more negative after May 2022 in central Stockholm, consistent with H3. This is in line with the broader geographic pattern documented in Section 5.1, where buyers in central Stockholm appear more likely to engage with TOA financial information. Across the remainder of the Swedish cooperative housing market, however, the rate shock appears to have reinforced rather than reduced fee anchoring, without improving the accuracy of the information the fee conveys.

5.3 Accounting quality and the limits of transparency

The accounting quality results reported in H4 and H5 are best understood in light of the fee anchoring mechanism established in the preceding sections. The H4 results show no statistically significant standalone association between accounting quality and transaction prices in the pooled sample (Eq. 6), and this null result holds across all three major counties and their surrounding municipalities. The only exception is Stockholm municipality, where the coefficient on AQ is suggested to be positive and marginally significant. At first glance, the near-universal null result might suggest that transparency is irrelevant in this market. We argue, however, that this interpretation is likely too strong, and that the results instead point to a conclusion about the conditions under which accounting quality may affect prices in a market populated exclusively by private buyers.

The distribution of AQ scores in the estimation sample (see Table 4.1) is heavily concentrated at the upper end of the scale. The median and first quartile of the aggregate AQ score are both 6, the maximum possible value, while the distribution indicates that approximately 87 per cent of all observations receive the maximum score, reflecting limited dispersion across the sample. This concentration of scores near the ceiling implies that the empirical tests of H4 and H5 are conducted with relatively little variation in the key variable of interest, which reduces statistical power and limits the ability to detect an accounting quality effect even if one exists in the underlying population. The null result for H4 and the limited geographic reach of the H5 result should therefore be interpreted with this constraint in mind.

With this caveat in mind, the pattern across hypotheses nonetheless introduces an important nuance. Stockholm municipality is the only submarket where three results point in the same direction where net debt is independently capitalised into prices (H2), accounting quality is positively associated with prices (H4), and accounting quality moderates the debt-price slope (H5). The convergence of these three findings in a single submarket is more informative than any of them in isolation, and is consistent with buyers in Stockholm engaging with TOA accounting information to a greater extent than buyers elsewhere. This interpretation finds support in the evidence of Berg (2026). When complex financial information was translated into an accessible rating displayed alongside price and size on Hemnet, apartment prices adjusted around rating thresholds in a way they had not before. The mechanism is the same one our results point to where financial information appears to reach prices when it is made visible at the point of decision, and to fail to do so when it requires buyers to actively seek it out. By extension, this suggests that making association debt more salient at the point of sale, rather than embedded in annual reports, could plausibly strengthen the debt capitalisation our results find largely absent outside Stockholm.

Our findings point in the same direction. Theory predicts that accounting quality should reduce information asymmetry and improve pricing accuracy through two channels, by increasing buyer capacity to interpret financial statements, and by making relevant information more accessible at the point of decision (Diamond and Verrecchia, 1991; Bordalo, Gennaioli and Shleifer, 2013). The Stockholm municipality results, where accounting quality carries a marginally significant price premium and moderates the debt-price slope in the predicted direction, suggest that buyers in this market possess sufficient sophistication to engage with and act on financial reporting quality. In the remaining markets, however, accounting quality does not appear to reach prices, suggesting that buyer sophistication may be insufficient to translate reporting quality into more accurate price formation.

Taken together, the geographic results suggest that fee anchoring is not uniform across markets. Stockholm municipality is the only submarket where buyers consistently appear to look beyond the monthly fee and incorporate underlying TOA financial information into prices. In Gothenburg, Malmö, and the surrounding county markets, debt appears to reach prices primarily through the monthly fee, while accounting quality does not systematically alter pricing behaviour. This reinforces the interpretation that the limiting factor is not only disclosure quality, but also buyer sophistication.

6. Conclusions

6.1 Concluding remarks

This paper set out to examine whether buyers in the Swedish cooperative housing market incorporate TOA financial fundamentals into apartment transaction prices, or whether they instead rely on the monthly fee as a heuristic proxy for total ownership cost, and further whether accounting quality and geographic buyer composition moderate this relationship. To study these questions, we construct a temporally matched dataset of over 200,000 apartment transactions linked to annual reports from more than 17,000 unique TOAs, and test five hypotheses within a hedonic pricing framework with fine-grained location and temporal fixed effects, estimated both on the full national sample and across regional subsamples.

On a national level the evidence tells the same story. Results indicate that monthly fees are negatively capitalised into prices, supporting H1, while TOA net debt carries no independent negative association with prices once the fee is controlled for, providing no support for H2. Rather than debt becoming more negatively priced after the Riksbank began tightening in 2022, the conditional association moved in a positive direction, providing no support for H3. Our analysis indicates that higher accounting quality does not carry a standalone price premium in the pooled sample, providing no support for H4, though it moderates the debt-price slope in the predicted direction, partially supporting H5. Stockholm municipality is a consistent exception, where H2, H3, H4, and H5 all receive directional support.

Taken together, these findings suggest that for most buyers in the Swedish cooperative housing market, the monthly fee functions as a sufficient statistic for total ownership cost, crowding out the balance sheet information that a more complete financial assessment would require. Stockholm municipality is the informative contrast. There, debt is independently capitalised into prices, accounting quality carries a marginal price premium, and accounting quality moderates the debt-price slope, suggesting that buyers possess sufficient sophistication to engage with and act on financial reporting quality. In the remaining markets, however, accounting quality does not appear to reach prices, the monthly fee remains the dominant financial signal, and even sharp increases in the cost of association debt do not appear sufficient to redirect buyer attention towards the balance sheet.

The findings have practical implications for both buyers and policymakers. Private buyers in the Swedish cooperative housing market may systematically underestimate the financial risk they take on when purchasing apartments in highly leveraged associations, with potentially significant consequences for household finances during periods of rising interest rates. If the salience mechanism our results point to also applies to net debt, interventions that make association net debt more visible at the point of sale could plausibly help close this gap without requiring buyers to become financial analysts. The Norwegian arrangement, where the buyer’s proportional share of association debt must be disclosed directly in property listings, offers one illustration of what such an intervention could look like in practice, though whether it would have the predicted effect in the Swedish market remains an open question.

In a market where the monthly fee is salient to most buyers and the balance sheet is not, buyers gain cognitive ease at the expense of financial accuracy. This may be the price of simplicity.

6.2 Contributions

We identify four main contributions of this paper. First, the thesis contributes methodologically to accounting research in decentralised, non-standardised source environments. The AI-assisted collection and extraction design has been a key enabler in this housing setting, where data is located in annual reports dispersed across thousands of sites and varies substantially in structure and content, and where conventional approaches would not be feasible at this scale. Although applied to Swedish TOAs, the design is relevant for empirical settings in which researchers need to locate data from unstructured sources and transform it into analysis-ready structured data, both within the accounting realm and beyond.

Second, it contributes to the literature on disclosure quality and the value relevance of accounting information by examining how financial disclosures are reflected in the Swedish cooperative apartment market. Prior Swedish work on TOA financial reporting quality has examined both reporting quality and apartment pricing, most notably Bokvist and Lanner (2018), whose analysis is restricted to the 100 largest associations. However, their pricing analysis does not directly test whether variation in accounting quality itself is associated with transaction prices. In contrast, our analysis links a disclosure-based accounting quality measure directly to apartment transactions across 17,703 associations, extending the disclosure-quality dimension from a narrower institutional setting to a broader market-level analysis of Swedish cooperative housing.

Third, the thesis contributes empirically by extending the analysis from a single market view to a structured decomposition across multiple submarkets of the Swedish cooperative apartment market. To our knowledge, prior research in a Swedish setting has either relied on large nationwide pooled regressions, which deliver average effects but conceal local variation, or restricted attention to a single geography, which delivers depth in one market but cannot speak to how pricing differs elsewhere. By examining both the pooled national level and the within-region structure this thesis identifies distinctions in how buyers price TOA financial information that have previously gone unobserved. This makes a notable contribution to the understanding of market heterogeneity that has been left out of earlier work, and enables the identification of regional differences that prior research has been unable to detect.

Fourth, the thesis contributes to the measurement of TOA financial position in hedonic valuation by refining how the asset side of the association balance sheet is treated. A notable component of association value, namely the non-owner area consisting of rental apartments and commercial premises, has to our knowledge been left unobserved in prior hedonic research on the Swedish cooperative market (Hjalmarsson & Hjalmarsson, 2006; Meyer & Ulmgren, 2018). Earlier studies have focused almost exclusively on fees and debt. Drawing on the operating-financial decomposition of Feltham and Ohlson (1995), the thesis incorporates non-owner area as a measure of operating-side assets that contribute positively to apartment value beyond apartment characteristics and net debt. Bringing this previously unobserved asset into the hedonic specification yields a more complete picture of the operating-side determinants of apartment value, particularly for the subset of associations where such space is material, and reduces the risk that variation in operating capacity is mistakenly attributed to V(ND) in designs that omit V(ONA) altogether.

6.3 Risks and limitations

Several limitations affect this study, spanning data construction, source coverage, and empirical design. The validation, correction, robustness, and archival procedures applied throughout may reduce but cannot eliminate their potential influence on the estimates.

First, the annual-report dataset captures only those TOAs whose reports were publicly discoverable through name-based web search at the time of collection, not the full universe of association filings. Associations without dedicated websites, those hosted on third-party platforms not surfaced by the search engine, and those with crawler-inaccessible navigation paths are therefore likely absent. As discovered in the PDF fetching stage, older filings are underrepresented (see Appendix A.5). One could also think that associations with more professionally maintained web presences are likely overrepresented. The collection process may also be unstable over time as websites may be redesigned, relocated, or taken offline between collection and any replication attempt, so exact reproduction of the crawling stage cannot be guaranteed. A separate concern is that LLM-based extraction may introduce field-level errors that persist despite manual validation. Relatedly, the LLM-based ranking for deeper crawling may introduce selection error if the model selected an incorrect or incomplete candidate site. As with all LLMs, the risk of unobservable biases in the training data is also seen as a risk.

Second, the Booli transaction data are provided “as is,” without guarantees of accuracy or completeness. Any omissions or measurement errors in the source propagate into the estimation sample independently of the annual-report pipeline.

Third, the empirical design relies on linear hedonic pricing models applied to a highly heterogeneous housing sample. Apartment prices are shaped by numerous dwelling, association, building, and neighbourhood characteristics, some of which are difficult to observe or measure consistently at scale. Although the regressions include detailed controls and fixed effects, the estimated coefficients should be interpreted as conditional associations rather than causal effects. The linear specification further imposes a common marginal relationship between each explanatory variable and log price, which may be restrictive if buyers respond differently to fees, debt, or accounting quality across price segments, regions, building vintages, or market conditions. The reported estimates therefore describe average pricing patterns in the observed sample and may mask nonlinearities and heterogeneous effects across submarkets.

Fourth, while our manual review of the randomized 30+3+3[12] sample provides support for the observed skewed distribution in accounting-quality scores to be rightful, it cannot rule out that the full-sample distribution is partly shaped by systematic scoring error from the AI model. This may affect the regression estimates and should therefore be considered when interpreting the results of regressions with AQ scoring.

Finally, the regressions cannot fully account for unobserved apartment-specific quality such as renovation standard, interior condition, and finishings. These features affect transaction prices but are not captured by the structural controls available in the data. Because interior renovation decisions are typically made by individual members rather than by the association, this variation is not expected to be mechanically related to association-level financial characteristics. However, if apartment quality is systematically correlated with association finances or with neighbourhood characteristics not fully absorbed by the controls, the estimated coefficients may still be affected. This limitation should therefore be read as a potential source of residual variation rather than one that can be ruled out entirely.

6.4 Suggestion for future research

Future research could extend the regional interpretation developed in this paper by incorporating local demographic and socioeconomic controls, such as income, education, age structure, household composition, and wealth indicators. The present study uses granular location fixed effects to account for local differences in neighbourhood quality, amenities, and socioeconomic composition, but these controls only approximate such variation indirectly. Since international hedonic housing-price research commonly includes neighbourhood socioeconomic characteristics and local amenities alongside property-level controls (Zabel, 2015), such data would provide a useful complement to the current specification. In particular, it would allow future studies to test more directly whether the stronger debt capitalisation and AQ moderation observed in Stockholm reflect differences in buyer sophistication and information-processing capacity, rather than broader unobserved differences across local housing markets.

Given the low variability of the AI-constructed AQ variable, where approximately 87% of observations receive the maximum score, further research could also develop the framework to capture more granular variation among high-scoring cases.

Future research could extend the analysis beyond the linear OLS framework used in this study by applying nonlinear or machine-learning-based models. Such approaches could help identify threshold effects, interaction patterns, and heterogeneous pricing dynamics across regions, building types, and market segments. They may also support more detailed analysis of smaller regional markets not separately examined in this report.

[12] The notation 30+3+3 refers to the initial randomised sample of 30 reports, plus three additional reports scoring 0 and three additional reports scoring 1 (see scoring validation procedure in Section 3.2.2)

7. References

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers. *Organizational Research Methods*, *16*(2), 270–301. https://doi.org/10.1177/1094428112470848

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Årsredovisningslag [Annual Accounts Act] (SFS 1995:1554). (1995). Sveriges Riksdag. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/arsredovisningslag-19951554_sfs-1995-1554/

Artemova, E., Tsvigun, A., Schlechtweg, D., Fedorova, N., Tilga, S., Chernyshev, K., & Obmoroshev, B. (2025). *Hands-On Tutorial: Labeling with LLM and Human-in-the-Loop*. Arxiv.org. https://arxiv.org/html/2411.04637v3

Atanasov, V., & Black, B. (2016). Shock-Based Causal Inference in Corporate Finance and Accounting Research. *Critical Finance Review*, *5*(2), 207–304. https://doi.org/10.1561/104.00000036

Barth, M. E., Beaver, W. H., & Landsman, W. R. (2001). The relevance of the value relevance literature for financial accounting standard setting: another view. *Journal of Accounting and Economics*, *31*(1-3), 77–104. https://doi.org/10.1016/s0165-4101(01)00019-2

Beattie, V., McInnes, B., & Fearnley, S. (2004). A methodology for analysing and evaluating narratives in annual reports: a comprehensive descriptive profile and metrics for disclosure quality attributes. *Accounting Forum*, *28*(3), 205–236. https://doi.org/10.1016/j.accfor.2004.07.001

Beretta, S., & Bozzolan, S. (2004). A framework for the analysis of firm risk communication. *The International Journal of Accounting*, *39*(3), 265–288. https://doi.org/10.1016/j.intacc.2004.06.006

Berg, P. (2026). Can ratings mitigate consumer inattention? Evidence from the Swedish housing market. *Journal of Economic Behavior & Organization*, *241*, 107349. https://doi.org/10.1016/j.jebo.2025.107349

Beyer, A., Cohen, D. A., Lys, T. Z., & Walther, B. R. (2010). The financial reporting environment: Review of the recent literature. *Journal of Accounting and Economics*, *50*(2-3), 296–343. https://doi.org/10.1016/j.jacceco.2010.10.003

Biddle, G. C., Hilary, G., & Verdi, R. S. (2009). How Does Financial Reporting Quality Relate to Investment efficiency? *Journal of Accounting and Economics*, *48*(2-3), 112–131. https://doi.org/10.1016/j.jacceco.2009.09.001

Bokföringsnämnden. (2023). *Vägledning: Årsredovisning i bostadsrättsförening.* BFN. https://www.bfn.se/wp-content/uploads/vl23-1-brf.pdf

Bokvist, K., & Lanner, J. (2018). *Accounting and Apartment Prices Evidence from the Swedish housing market*. https://arc.hhs.se/download.aspx?MediumId=4029

Bordalo, P., Gennaioli, N., & Shleifer, A. (2013). Salience and Consumer Choice. *Journal of Political Economy*, *121*(5), 803–843. https://doi.org/10.1086/673885

Bostadsrätterna. (2022, December 9). *Ändrad beräkningsgrund för oäkta föreningar*. Bostadsrätterna | Nyheter. https://www.bostadsratterna.se/nyheter/artiklar/2022/aendrad-berakningsgrund-for-oakta-foreningar

Bostadsrättsägarna. (2024). *Nya regler för årsredovisning och revisionsberättelse - Bostadsrättsägarna*. Bostadsrattsagarna.com. https://bostadsrattsagarna.com/press-media/nyhetsarkiv/284-nya-regler-foer-arsredovisning-och-revisionsberaettelse

Bostadsrättslag [Housing Cooperative Act] (SFS 1991:614). (1991). Sveriges Riksdag. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/bostadsrattslag-1991614_sfs-1991-614/

Botosan, C. A. (1997). Disclosure Level and the Cost of Equity Capital. *The Accounting Review*, *72*(3), 323–349. JSTOR. https://doi.org/10.2307/248475

Boverket. (2022). *Nyckeltal ska redovisas i den ekonomiska planen*. Boverket. https://www.boverket.se/sv/ekonomiska-planer/ekonomisk-plan/nyckeltal/

Bushman, R. M., & Smith, A. J. (2001). Financial Accounting Information and Corporate Governance. *Journal of Accounting and Economics*, *32*(1-3), 237–333. https://doi.org/10.1016/S0165-4101(01)00027-1

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

Capozza, D. R., & Helsley, R. W. (1989). The fundamentals of land prices and urban growth. *Journal of Urban Economics*, *26*(3), 295–306. https://doi.org/10.1016/0094-1190(89)90003-x

Case, K. E., & Shiller, R. J. (1988, February 1). *The Efficiency of the Market for Single-Family Homes*. National Bureau of Economic Research. https://www.nber.org/papers/w2506

Correia, S. (2016). *A Feasible Estimator for Linear Models with Multi-Way Fixed Effects*. https://scorreia.com/research/hdfe.pdf

Dagdelen, J., Dunn, A., Lee, S., Walker, N., Rosen, A. S., Ceder, G., Persson, K. A., & Jain, A. (2024). Structured information extraction from scientific text with large language models. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-45563-x

Diamond, D. W., & Verrecchia, R. E. (1991). Disclosure, Liquidity, and the Cost of Capital. *The Journal of Finance*, *46*(4), 1325–1359. https://doi.org/10.1111/j.1540-6261.1991.tb04620.x

Du, Y., Tian, M., Ronanki, S., Rongali, S., Bodapati, S., Galstyan, A., Wells, A., Schwartz, R., Huerta, E. A., & Peng, H. (2025). *Context Length Alone Hurts LLM Performance Despite Perfect Retrieval*. ArXiv.org. https://arxiv.org/abs/2510.05381

Eiendomsmeglingsloven. (2007, June 29). *Lov om eiendomsmegling (LOV-2007-06-29-73)*. Lovdata. https://lovdata.no/dokument/NL/lov/2007-06-29-73

Erik S., & Zhang, B. (2024, December 19). *Building effective agents*. Anthropic.com. https://www.anthropic.com/research/building-effective-agents

Fastighetsägarna. (2017). *Vad är en äkta och oäkta förening? - Fastighetsägarna*. https://www.fastighetsagarna.se/aktuellt/nyheter/fastighetsagarna-service/vad-ar-akta-och-oakta-foreningar/

Feltham, G. A., & Ohlson, J. A. (1995). Valuation and Clean Surplus Accounting for Operating and Financial Activities. *Contemporary Accounting Research*, *11*(2), 689–731. https://doi.org/10.1111/j.1911-3846.1995.tb00462.x

Ferguson, N., Pennington, J., Beghian, N., Mohan, A., Kiela, D., Agrawal, S., & Nguyen, T. H. (2026). *ExtractBench: A Benchmark and Evaluation Methodology for Complex Structured Extraction*. ArXiv.org. https://arxiv.org/abs/2602.12247

FI. (2010). 2010:2 | Finansinspektionen. FI. https://www.fi.se/sv/vara-register/fffs/sok-fffs/2010/20102/

Francis, J., LaFond, R., Olsson, P., & Schipper, K. (2005). The market pricing of accruals quality. *Journal of Accounting and Economics*, *39*(2), 295–327. https://doi.org/10.1016/j.jacceco.2004.06.003

Fredrik Fexe, & Berglind, A. W. (2026, March 25). *Tillfällig räntelättnad för bostadsrättsföreningar men höjningar väntar runt hörnet*. Mynewsdesk; Nabo. https://www.mynewsdesk.com/se/nabo-group-ab/pressreleases/tillfaellig-raentelaettnad-foer-bostadsraettsfoereningar-men-hoejningar-vaentar-runt-hoernet-3439518

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(30). https://doi.org/10.1073/pnas.2305016120

Goodman, A. C., & Thibodeau, T. G. (1998). Housing Market Segmentation. *Journal of Housing Economics*, *7*(2), 121–143. https://doi.org/10.1006/jhec.1998.0229

Gutierrez-Bustamante, M., & Espinosa-Leal, L. (2022). Natural Language Processing Methods for Scoring Sustainability Reports - A Study of Nordic Listed Companies. *Sustainability*, *14*(15), 9165. https://doi.org/10.3390/su14159165

Herk, N. (2026). WAT CLAUDE.md - Claude Code · AI Automation Society. Skool.com. https://www.skool.com/ai-automation-society/classroom/076a1c6e?md=122d76fe19984887af89e30ba0c7d2f8

Heyman, A. V., & Sommervoll, D. E. (2019). House prices and relative location. *Cities*, *95*, 102373. https://doi.org/10.1016/j.cities.2019.06.004

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1-3), 337–386. https://doi.org/10.1016/j.jacceco.2003.10.002

Hjalmarsson, E., & Hjalmarsson, R. (2006). *Efficiency in Housing Markets: Do Home Buyers Know How to Discount? International Finance Discussion Papers Board of Governors of the Federal Reserve System*.

HSB. (2021, October 14). *Juridisk person som ägare av bostadsrättslägenhet – det här gäller*. HSB.se. https://www.hsb.se/nyheter-och-tips/kunskapsbank/juridisk-person-som-agare-av-bostadsrattslagenhet--det-har-galler/

Jansen, T., Liebenow, L. W., Schaller, N.-J., Hattie, J., & Möller, J. (2026). Automated Data Extraction by Large Language Models: Assessing Accuracy in Comparison to Human

Experts Using the Example of Visible Learning. *Educational Psychology Review*, *38*(1). https://doi.org/10.1007/s10648-026-10136-5
Kahneman, D. (2011). *Thinking, fast and slow*. Psycnet.apa.org. https://psycnet.apa.org/record/2011-26535-000
Kutner, M. H., Nachtsheim, C., & Neter, J. (2018). *Applied linear regression models* (4th ed.). Mcgraw-Hill Education/Asia.
Kwak, S. G., & Kim, J. H. (2017). Central Limit theorem: the Cornerstone of Modern Statistics. *Korean Journal of Anesthesiology*, *70*(2), 144–156. https://doi.org/10.4097/kjae.2017.70.2.144
Lag om ekonomiska föreningar [Economic Associations Act] (SFS 2018:672). (2018). Sveriges Riksdag. https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-2018672-om-ekonomiska-foreningar_sfs-2018-672/
Li, Y., Ramprasad, R., & Zhang, C. (2024). *A Simple but Effective Approach to Improve Structured Language Model Output for Information Extraction* (pp. 5133–5148). https://aclanthology.org/2024.findings-emnlp.295.pdf
Liu, S., Liu, Y., Wang, Z., Wang, Y., Wu, H., Xiang, L., & He, Z. (2025). *Select-Then-Decompose: From Empirical Analysis to Adaptive Selection Strategy for Task Decomposition in Large Language Models*. Arxiv.org. https://arxiv.org/html/2510.17922v1#bib.bib44
Lusardi, A., & Mitchell, O. S. (2014). The Economic Importance of Financial Literacy: Theory and Evidence. *Journal of Economic Literature*, *52*(1), 5–44. https://doi.org/10.1257/jel.52.1.5
Marston, C. L., & Shrives, P. J. (1991). The use of disclosure indices in accounting research: A review article. *The British Accounting Review*, *23*(3), 195–210. https://doi.org/10.1016/0890-8389(91)90080-l
Meyer, W., & Ulmgren, M. (2018). *A ticking bomb? Leverage in Swedish Housing cooperatives*. Bachelor thesis in Finance. Stockholm: Stockholm School of Economics.
Mize, T. (2019). Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects. *Sociological Science*, *6*, 81–117. https://doi.org/10.15195/v6.a4
Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review*, *48*(3), 261–297. https://www.jstor.org/stable/1809766?seq=1
Nabo. (2025). *Belåning -Sveriges bostadsrättsföreningar*. https://nabo.se/wp-content/uploads/2025/02/Nabo_Belaning_Sveriges_Bostadsrattsforeningar_2025.pdf
Narwold, A., Sah, V., & Conroy, S. J. (2018). Impact of Homeowners Association Fees on Condominium Prices. *Journal of Housing Research*, *27*(1), 79–91. https://doi.org/10.1080/10835547.2018.12092142
Osborne, J., & Overbay, A. (2004). The power of outliers (and why researchers should ALWAYS check for them). *Practical Assessment, Research, and Evaluation*, *9*(6), 6. https://doi.org/10.7275/qf69-7k43

Robertsen, K., & Theisen, T. (2011). The Impact of Financial Arrangements and Institutional Form on Housing Prices. *The Journal of Real Estate Finance and Economics*, *42*(3), 371–392. https://doi.org/10.1007/s11146-009-9213-z

SBC. (2023a, March 8). *Stambyte i bostadsrättsförening*. SBC; SBC Sveriges Bostadsrättscentrum AB. https://www.sbc.se/lar-dig-mer/styrelse/kunskapsartiklar/stambyte-i-bostadsrattsforening/

SBC. (2023b, March 31). *Det här gäller för brf:er med tomträtt*. SBC; SBC Sveriges Bostadsrättscentrum AB. https://www.sbc.se/lar-dig-mer/styrelse/kunskapsartiklar/tomtratt/

SCB. (2024, February 6). *Inkomster för personer*. Statistiska Centralbyrån. https://www.scb.se/hitta-statistik/sverige-i-siffror/utbildning-jobb-och-pengar/inkomster-for-personer/

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665–690. https://doi.org/10.1016/s0304-3932(03)00029-1

Skatteverket. (n.d.). *Beskattning av andel i oäkta bostadsföretag*. Retrieved March 15, 2026, from https://www.skatteverket.se/privat/fastigheterochbostad/beskattningavandelioaktabostadsforetag.4.6d02084411db6e252fe80009354.html

Solna Stad. (2026). *Tomträtt*. Solna.se. https://www.solna.se/bygga-bo--miljo/boende-miljo-och-natur/tomtratt

Stockholms Stad. (2026, February 17). *Tomträtt och arrende - Stockholms stad*. Boende.stockholm. https://boende.stockholm/regler-tillstand/tomtratt-och-arrende/

Sveriges Riksbank. (2026, May 13). *Policy rate (previously repo rate), deposit and lending rate*. Www.riksbank.se. https://www.riksbank.se/en-gb/statistics/interest-rates-and-exchange-rates/policy-rate-deposit-and-lending-rate/

Utlånsforskriften. (2020, December 9). *Forskrift om finansforetakenes utlånspraksis (FOR-2020-12-09-2648)*. Lovdata. https://lovdata.no/dokument/SF/forskrift/2020-12-09-2648

van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), 449–472. https://doi.org/10.1016/j.jfineco.2011.03.006

Wilhelmsson, M. (2008). House price depreciation rates and level of maintenance. *Journal of Housing Economics*, *17*(1), 88–101. https://doi.org/10.1016/j.jhe.2007.09.001

Woetzel, J., Mischke, J., Madgavkar, A., Windhagen, E., Smit, S., Birshan, M., Kemeny, S., & J. Anderson, R. (2021, November 15). *The rapid growth in global wealth | McKinsey*. McKinsey. https://www.mckinsey.com/industries/financial-services/our-insights/the-rise-and-rise-of-the-global-balance-sheet-how-productively-are-we-using-our-wealth

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach.* (5th ed.). Mason, OH: South-Western Cengage Learning.

Xiao, Y.-A., Gao, P., Peng, C., & Xiong, Y. (2026). *Reducing Cost of LLM Agents with Trajectory Reduction*. Arxiv.org. https://arxiv.org/pdf/2509.23586

Xie, Z., & Sun, Z. (2026). *Mitigating Hallucination on Hallucination in RAG via Ensemble Voting*. Arxiv.org. https://arxiv.org/abs/2603.27253

Zabel, J. (2015). The hedonic model and the housing cycle. *Regional Science and Urban Economics*, *54*, 74–86. https://doi.org/10.1016/j.regsciurbeco.2015.07.005

Zhang, W., Zeng, L., Xiao, Y., Li, Y., Cui, C., Zhao, Y., Hu, R., Liu, Y., Zhou, Y., & An, B. (2025). *AgentOrchestra: Orchestrating Hierarchical Multi-Agent Intelligence with the Tool-Environment-Agent(TEA) Protocol*. Arxiv.org. https://arxiv.org/html/2506.12508v4

Appendix

A.1. Workflow file (.md)

Agent Instructions

You're working inside the **WAT framework** (Workflows, Agents, Tools). This architecture separates concerns so that probabilistic AI handles reasoning while deterministic code handles execution. That separation is what makes this system reliable.

The WAT Architecture

Layer 1: Workflows (The Instructions)

- Markdown SOPs stored in `workflows/`

- Each workflow defines the objective, required inputs, which tools to use, expected outputs, and how to handle edge cases

- Written in plain language, the same way you'd brief someone on your team

Layer 2: Agents (The Decision-Maker)

- This is your role. You're responsible for intelligent coordination.

- Read the relevant workflow, run tools in the correct sequence, handle failures gracefully, and ask clarifying questions when needed

- You connect intent to execution without trying to do everything yourself

- Example: If you need to pull data from a website, don't attempt it directly. Read `workflows/scrape_website.md`, figure out the required inputs, then execute `tools/scrape_single_site.py`

Layer 3: Tools (The Execution)

- Python scripts in `tools/` that do the actual work

- API calls, data transformations, file operations, database queries

- Credentials and API keys are stored in `.env`

- These scripts are consistent, testable, and fast

Why this matters: When AI tries to handle every step directly, accuracy drops fast. If each step is 90% accurate, you're down to 59% success after just five steps. By offloading execution to deterministic scripts, you stay focused on orchestration and decision-making where you excel.

How to Operate

1. Look for existing tools first

Before building anything new, check `tools/` based on what your workflow requires. Only create new scripts when nothing exists for that task.

2. Learn and adapt when things fail

When you hit an error:

- Read the full error message and trace

- Fix the script and retest (if it uses paid API calls or credits, check with me before running again)

- Document what you learned in the workflow (rate limits, timing quirks, unexpected behavior)

- Example: You get rate-limited on an API, so you dig into the docs, discover a batch endpoint, refactor the tool to use it, verify it works, then update the workflow so this never happens again

3. Keep workflows current

Workflows should evolve as you learn. When you find better methods, discover constraints, or encounter recurring issues, update the workflow. That said, don't create or overwrite workflows without asking unless I explicitly tell you to. These are your instructions and need to be preserved and refined, not tossed after one use.

The Self-Improvement Loop

Every failure is a chance to make the system stronger:

1. Identify what broke

2. Fix the tool

3. Verify the fix works

4. Update the workflow with the new approach

5. Move on with a more robust system

This loop is how the framework improves over time.

File Structure

What goes where:

- **Deliverables**: Final outputs go to cloud services (Google Sheets, Slides, etc.) where I can access them directly

- **Intermediates**: Temporary processing files that can be regenerated

Directory layout:

```
.tmp/           # Temporary files (scraped data, intermediate exports). Regenerated as needed.
tools/          # Python scripts for deterministic execution
workflows/      # Markdown SOPs defining what to do and how
.env            # API keys and environment variables (NEVER store secrets anywhere else)
credentials.json, token.json  # Google OAuth (gitignored)
```

Core principle: Local files are just for processing. Anything I need to see or use lives in cloud services. Everything in `.tmp/` is disposable.

Bottom Line

You sit between what I want (workflows) and what actually gets done (tools). Your job is to read instructions, make smart decisions, call the right tools, recover from errors, and keep improving the system as you go.

Stay pragmatic. Stay reliable. Keep learning.

Notes: This Markdown file, taking the same shape as Herk (2026), provides the initial structure used by the agent to create its development environment. Subsequent human-supported agentic runs of the workflow adapt this file for the purpose of this paper. It is specifically designed for the three-layer design used in this paper and is included as a pragmatic design choice rather than as an academically validated or optimally aligned prompt.

A.2. Candidate selection prompt

You are an expert web-search evaluator.

You are given JSON search results for the association [association name].

Your task is to select the best result to scrape next.

Selection criteria:

1. Prefer the association's official website.
    a. Example: a website such as "https://www.brfbrommahusen.se/" is likely to be official because the domain closely matches the association name BRF Brommahusen.
2. If no official website is present, select the result most likely to contain annual reports for the association.
3. Avoid general directory sites, maps/listing services, and social media pages unless no better option exists.

Return the position number of the selected result from the JSON search results.

Note: The prompts have been translated from Swedish using ChatGPT 5.5. The association name is dynamically inserted for each query. "BRF Brommahusen" is a fictitious example used to clarify the task. The model output is constrained to a single number between 1 and 10, reflecting the ten candidate results returned by Firecrawl. After excluding sites that should not be scraped and sites handled by a separate pipeline (see discussion in 3.2.2), the candidate list provided to the AI may contain fewer than ten results.

A.3. Screening, extraction and scoring prompts per AI call

Table A.3.1. Prompts related to first call

(Call 1) **Part**	**Prompt**
Classification	Step 1 - Classification: Determine whether the PDF file is a complete annual report for a housing cooperative / tenant-owner association, specifically a Swedish bostadsrättsförening, and not a limited company, foundation, or any other legal form. Set is_annual_report to true only if the PDF clearly is a complete annual report for a bostadsrättsförening. Set is_annual_report to false for bylaws, energy declarations, prospectuses, economic plans, appendices, or incomplete documents. If is_annual_report is false, all other fields must be returned as null.
Extraction	Step 2 - Extraction: If is_annual_report is true, extract the remaining fields according to the schema. If is_annual_report is false, leave all remaining fields as null. General rules: - Use only information that is stated in the document. - Never guess. Return null if the information is not sufficiently clear. - Do not derive your own calculations. - If the same information appears in multiple places, use the most explicit information from the main sections of the annual report. - If the information is contradictory, return null. - Do not use information that appears only in the auditor's report, table of contents, or standalone appendices. - Keep negative values as negative. - All monetary amounts must be returned in Swedish kronor. - If amounts are stated in thousands of SEK or millions of SEK, convert them to SEK. - If the unit is not clearly stated, return null for that field. - If no cash flow statement exists: set existerar_kassaflodesanalys = false and all other cash flow fields to null. BRF-specific interpretation rules: - Residential apartments should be interpreted as tenant-owned apartments unless something clearly indicates that they refer to rental apartments. - lagenhetsyta_kvm may be set from information about apartment area, living area, or residential area for the association's residential apartments. - antal_bostadsratter may be set from the number of residential apartments if nothing indicates that they are rental apartments. - akta_forening: true if the association is explicitly stated to be a genuine housing cooperative / private housing company; false if the opposite is explicitly stated; otherwise null. - tomtratt: true if leasehold land is mentioned or a leasehold fee / ground rent is reported. - tomtratt: false if the land is clearly owned. Otherwise, return null for tomtratt.

Note: Prompts have been translated from Swedish using ChatGPT 5.5. In practice, these prompts are combined with prompt schema in Table A.3.2. to define the variables of interest.

Table A.3.2. Prompt schema for AI extraction.

Variable name	Prompt
is_annual_report	True only if the PDF is clearly a complete annual report for a housing cooperative (bostadsrättsförening). False for bylaws, energy declarations, prospectuses, financial plans, appendices, or incomplete documents.
namn_pa_forening	Full name of the association exactly or near-exactly as it appears in the annual report.
organisationsnummer	The association's registration number, preferably formatted with a hyphen, e.g. 716409-5874. Null if not clearly stated.
rakenskapsar_start	First date of the fiscal year in ISO format YYYY-MM-DD. Null if not clearly stated.
rakenskapsar_slut	Last date of the fiscal year in ISO format YYYY-MM-DD. Null if not clearly stated.
ar	Calendar year of rakenskapsar_slut, e.g. 2024 if end date is 2024-12-31. Null if not clearly stated.
lagenhetsyta_kvm	Residential apartment area in m² for owner-occupied units (bostadsrätt). If rental units are not separately mentioned, the full apartment area is assumed to be owner-occupied. Null if data is missing.
lagenhetsyta_hyresratt_kvm	Rental apartment area in m². Null if missing or if no rental units are mentioned.
lokalyta_kvm	Commercial/premises area in m². Null if not found.
akta_forening	True only if the association is explicitly stated to be a genuine housing cooperative (äkta bostadsrättsförening / privatbostadsföretag). False only if the opposite is explicitly stated. Otherwise null.
tomtratt	True if any of the following apply: site leasehold (tomträtt) is explicitly mentioned; a ground rent fee is reported in the income statement or notes; there is a description of a rent review period or municipal land agreement indicating the land is not owned. False if no leasehold indicators exist anywhere in the document. Null if the document cannot be read.
summa_kassa_och_bank	Cash and bank balance, or a clearly equivalent total in the balance sheet. Amount in SEK.
langfristig_del_av_skulder_till_kreditinstitut	The portion of liabilities to credit institutions that falls due more than one year after the balance sheet date. Amount in SEK.
kortfristig_del_av_skulder_till_kreditinstitut	Short-term portion of liabilities to credit institutions. Can be found in the credit institution notes if not specified in the balance sheet. Amount in SEK.

Note: Prompts have been translated from Swedish using ChatGPT 5.5. The table displays a subset of variables from the full prompt, limited to those relevant for the purposes of this thesis.

Table A.3.3. Prompts related to second call

(Call 2) **Part**	**Prompt**
Quality scoring	You are a strict research assistant specialising in accounting. Assess the annual report across four dimensions using a 0-2 scale. Do not make assumptions without support in the text. If evidence is missing, choose the lower score. Dimension A - Future maintenance: - 0 = no or vague information. - 1 = measures are mentioned without timing, cost, financing, or fee impact. - 2 = clear measures with timing, cost, financing, or fee impact. Dimension B - Risk / financing: - 0 = no risks or only generic risks. - 1 = risk is mentioned without effect or action. - 2 = clear BRF-specific risks linked to finances, exposure, or action. Dimension C - Events: - 0 = no events or only standard text. - 1 = events are mentioned without clear consequence. - 2 = events are linked to result, financial position, liquidity, debt, or the future. Dimension U - Overall uncertainty: - 3 = clear report. - 2 = some uncertainty. - 1 = very vague or difficult-to-interpret report. For dimensions A, B, and C, return: - score: integer from 0 to 2 For U, return: - score: integer from 1 to 3

Note: Prompts have been translated from Swedish using ChatGPT 5.5

A.4. Accounting quality distribution statistics

Table A.4.1. AQ score distribution

AQ dimension	2	1	0
Future maintenance disclosure (A)	95.5%	3.4%	1.1%
Risk and financing disclosure (B)	91.5%	7.4%	1.1%
Important events disclosure (C)	89.4%	9.2%	1.4%
Average	**92.1%**	**6.7%**	**1.2%**

Note: Table shows the scoring (0-2) distribution across the accounting quality dimensions (A-C) where 2 denotes the highest accounting quality score and 0 the lowest.

Table A.4.2. AQ uncertainty distribution

AQ dimension	1 = low uncertainty / clear report	2 = moderate uncertainty / some ambiguity	3 = high uncertainty / vague or difficult-to-interpret report
A	97.3%	2.5%	0.2%
B	93.4%	6.5%	0.1%
C	91.3%	8.5%	0.2%
Average	**94.0%**	**5,8%**	**0.1%**

Note: Table shows accounting quality confidence scoring distribution across the three accounting quality dimensions. A = Future maintenance disclosure, B = Risk and financing disclosure, C = Important events disclosure.

Table A.4.3. Combined AQ score distribution (A+B+C)

AQ total	%
0	0.82%
1	0.16%
2	0.69%
3	2.40%
4	3.07%
5	5.44%
6	87.42%

Note: The table shows the distribution of the total accounting quality score, calculated as the sum of all accounting dimension scores.

A.5. Retention after temporal matching

Year	Org-linked	Temporal-linked	Retention
2012	6,284	470	7.5%
2013	26,775	2,343	8.8%
2014	31,560	3,496	11.1%
2015	37,183	4,825	13.0%
2016	41,768	5,653	13.5%
2017	46,619	6,156	13.2%
2018	52,162	7,796	14.9%
2019	65,238	13,616	20.9%
2020	95,699	28,589	29.9%
2021	105,819	43,357	41.0%
2022	89,870	34,820	38.7%
2023	76,356	29,119	38.1%
2024	84,637	42,493	50.2%
2025	81,911	56,614	69.1%
2026	13,874	11,660	84.0%
	855,755	291,007	

Note: The table reports the yearly distribution before and after temporal matching, with retention rates showing the percentage of observations retained. Retention rises with transaction year, reflecting that associations and third-party sites post recent filings more reliably than historical ones.

A.6. Control variable availability and evaluation

The following table reports the availability of potential control variables that are not central to the main hypotheses but were considered during model specification. Availability is calculated before the final regression sample restrictions and is defined as the number and share of non-missing observations out of 291,007 matched transactions. The table is used to motivate the choice of control control variables, which prioritise variables with high availability, theoretical relevance, and support in prior hedonic housing-price research.

Table A.6.1. Control variable availability and evaluation

Variable	Description	Available %	Control?	Reasoning
livingArea	Apartment living area	99.55%	Yes	Standard hedonic control
rooms	Number of rooms	99.68%	Yes	Standard hedonic control
floor	Apartment floor	88.84%	Yes	Common hedonic control
hasBalcony	Balcony indicator	62.10%	No	Low availability
listPrice	Listing price	92.93%	No	Potentially endogenous to transaction price
akta_forening	Qualified housing association indicator	99.04%	Yes	Controls for legal/tax classification
tomtratt	Ground lease indicator	98.05%	Yes	Controls for ownership/leasehold status
byggar	Building year	92.67%	Yes	Captures building age, but is converted (See Section 3.3.2)
hyresratt_lokal_area_ratio	Non-owner area ratio	99.26%	Yes	Controls for association asset structure and income potential
buildingHasElevator	Elevator indicator	33.91%	No	Low availability
hasFireplace	Fireplace indicator	34.34%	No	Low availability
hasPatio	Patio indicator	35.61%	No	Low availability
heating	Heating type	19.5%	No	Low availability
plotArea	Plot area	1.31%	No	Low availability
assessedValue	Assessed value	0%	No	Low availability

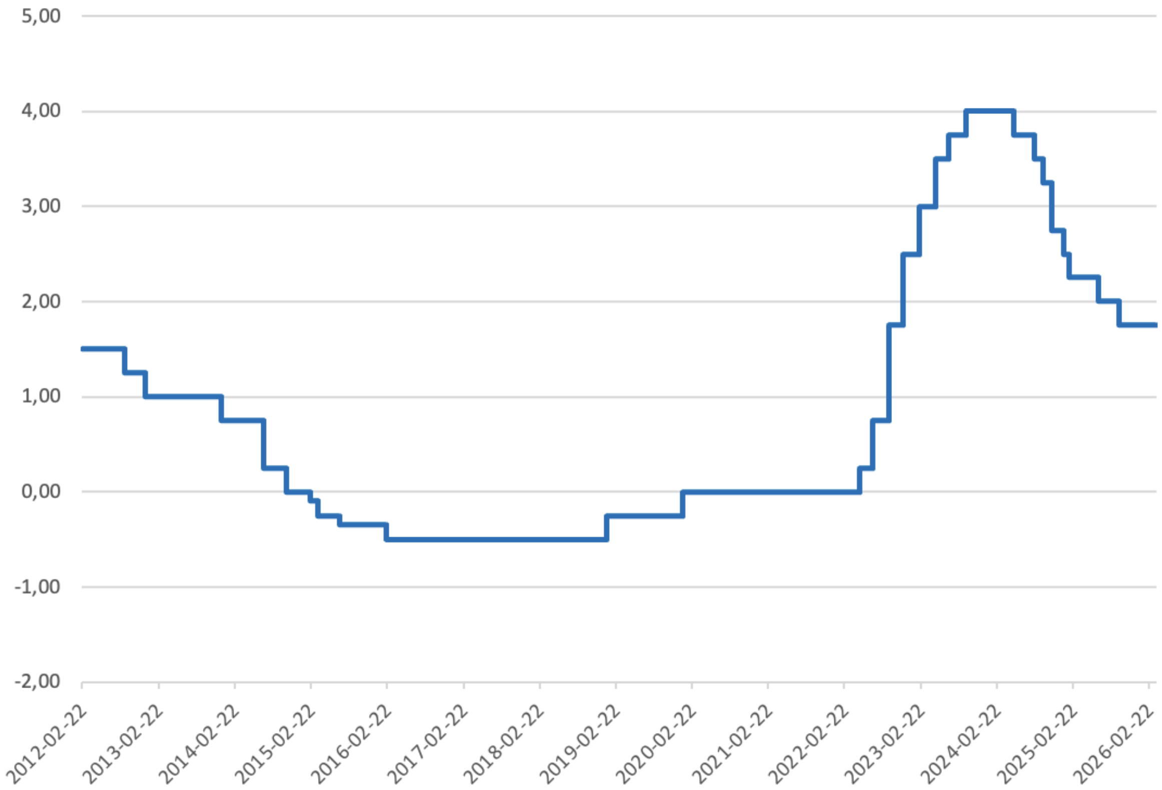
Notes: Variables related to fixed effects are excluded.

A.7. Construction-period indicators

Construction-period variable	Definition
pre_1934	built before 1934
1934_1957	built from 1934 to 1957
1958_1964	built from 1958 to 1964
1965_1974	built from 1965 to 1974
1975_1993	built from 1975 to 1993
1994_2009	built from 1994 to 2009
2010_plus	built in 2010 or later
unknown	missing construction year

Note: The construction-year controls follow Booli's established periodisation of Swedish housing construction years. These bins are intended to capture historically meaningful building-era differences in Swedish housing stock, including older stock, functionalist housing, the Folkhem period, the Million Programme era, post-subsidy construction, modern construction, and new construction. Missing construction years are retained as a separate unknown category rather than dropped.

A.8. Policy rate development during the study period



Source: Riksbanken.se

A.9. Robustness testing

The main results reported in Sections 4.2 and 4.3 are subjected to a series of robustness checks. The purpose of these checks is not to establish causality, but to assess whether the main patterns are sensitive to alternative measurement choices, sample restrictions, location fixed-effect definitions, influential observations, or alternative accounting-quality constructions. While numerous robustness tests could be conducted, we balanced potential impact against time constraints to focus on the most informative subset.

A.9.1. Robustness design

The first set of checks concerns sample construction and variable measurement. Since the main analysis applies a symmetric 1 per cent trim to Fee and Debt, the trim threshold is varied between 0.5 per cent and 2 per cent to assess whether the results depend on this particular choice. In addition, net debt is replaced with gross debt in Eq. (4) to examine whether the debt result is driven by the subtraction of cash holdings from total liabilities. Gross debt is a relevant alternative measure because current Swedish reporting rules for tenant-owned housing associations require debt per square metre to be reported as a standard key ratio, calculated using interest-bearing liabilities at the balance-sheet date (Bokföringsnämnden, 2023; Boverket, 2022). This measure is therefore closer to the debt metric that buyers may observe directly in annual reports.

The second set of checks concerns model specification and fixed effects. The main specifications use DeSO 2025 location fixed effects. As a sensitivity check, Eq. (2) is re-estimated using five-digit postcode fixed effects instead, which tests whether the fee-capitalisation result depends on the chosen spatial granularity. Since postcode cells contain fewer observations on average than DeSO cells (Table 3.3), the postcode specification leaves less within-cell variation available for identification, a trade-off that should be borne in mind when comparing the two estimates.

The third set addresses sample composition and influential observations. Although construction-period indicators control for average differences across building cohorts, they may not fully capture unobserved within-group quality differences such as recent renovation activity. Eq. (4) is therefore re-estimated on the subsample of apartments in TOAs constructed in 2010 or later. This restriction focuses on the newest construction bucket, where major structural renovations are less likely to have occurred, and assesses whether the main debt result is also present in a sample less exposed to renovation-related omitted-variable concerns. In addition, the top 0.1 per cent of TOAs by maximum Cook's distance are excluded to assess whether the main debt result is driven by a small number of influential associations.

The fourth set of checks concerns time-period interpretation. To assess whether the post-2022 debt interaction reflects the specific timing of the Riksbank rate-tightening cycle rather than a general difference between earlier and later periods, Eq. (5) is re-estimated within the pre-2022 subsample using placebo break dates on 4 May 2020 and 4 May 2021. This follows the logic of placebo tests in shock-based empirical designs, where researchers examine whether similar effects appear in periods when the hypothesised shock has not yet occurred (Atanasov and Black, 2016). Eq. (5) is also re-estimated excluding 2020-2021 to account for any bias related to Covid-years.

Finally, the accounting-quality results are examined using alternative AQ constructions. The H4 specification is decomposed into the three individual AQ components with H5 moderation specification re-estimated separately using each individual AQ component in the debt interaction. These checks assess whether the AQ results are driven by the composite index or by a particular disclosure dimension.

A.9.2. Robustness results

Overall, the robustness checks support the stability of the main findings, although they also indicate that the H2 results should be interpreted cautiously. As seen in Table A.9.1, the fee-capitalisation result is robust to using postcode rather than DeSO fixed effects, despite the reduced within-cell variation under the finer spatial specification, suggesting that the negative association between monthly fees and transaction prices is not specific to the baseline location fixed-effect definition. The debt results are also stable across alternative trimming thresholds and when gross debt is used instead of net debt, indicating that the positive direct association between TOA debt and price is not mainly driven by the chosen trim threshold or by the cash-netting procedure.

Table A.9.1. Robustness checks by specification

Robustness check	H1 Fee coefficient Eq. (2)	H2 Debt coefficient Eq. (4)	H3 Debt × Post2022 Eq. (5)	H4 AQ / component coefficient Eq. (6)	H5 Debt × AQ coefficient Eq. (7)
Baseline specification	−0.0035***	+2.2e−06***	+3.8e−06***	−0.0005	−6.4e−07***
Postcode FE	−0.0056***				
0.5% trim		+1.9e−06***	+3.3e−06***		
2% trim		+2.8e−06***	+4.2e−06***		
Gross debt		+2.0e−06***			
Only 2010+ buildings		−7.36e−07			
Drop top 0.1% Cook's distance TOAs		+2.3e−06***			
Placebo break: 2020			+1.0e−07		
Placebo break: 2021			+6.9e−07		
Exclude Covid			+4.5e−06***		
AQ_A component				−0.0028	
AQ_B component				+7.5e−05	
AQ_C component				−0.0010	
AQ_A-only interaction					−1.7e−06***
AQ_B-only interaction					−1.9e−06***
AQ_C-only interaction					−1.1e−06***

Notes: The table reports the main coefficient from each robustness specification. Empty cells indicate that the robustness check is not applicable to that hypothesis or equation.
*Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.*

When apartments in TOAs constructed in 2010 or later are looked at in isolation, where building quality should be more homogeneous, the national coefficient on Debt in Eq. (4) is negative but statistically insignificant. The estimate is -7.4×10^{-7}, compared with $+2.2 \times 10^{-6}$ in the full national sample.

The placebo break-date tests appear to support the interpretation that the post-2022 result is related to the actual rate-tightening period. When Eq. (5) is re-estimated within the pre-2022 subsample using placebo break dates in May 2020 and May 2021, the interaction coefficients are $+1.0 \times 10^{-7}$ and $+6.9 \times 10^{-7}$ respectively, both an order of magnitude smaller than the baseline post-2022 interaction of $+3.8 \times 10^{-6}$ and neither statistically significant. This suggests that the main interaction is not simply a mechanical result of splitting the sample into earlier and later periods.

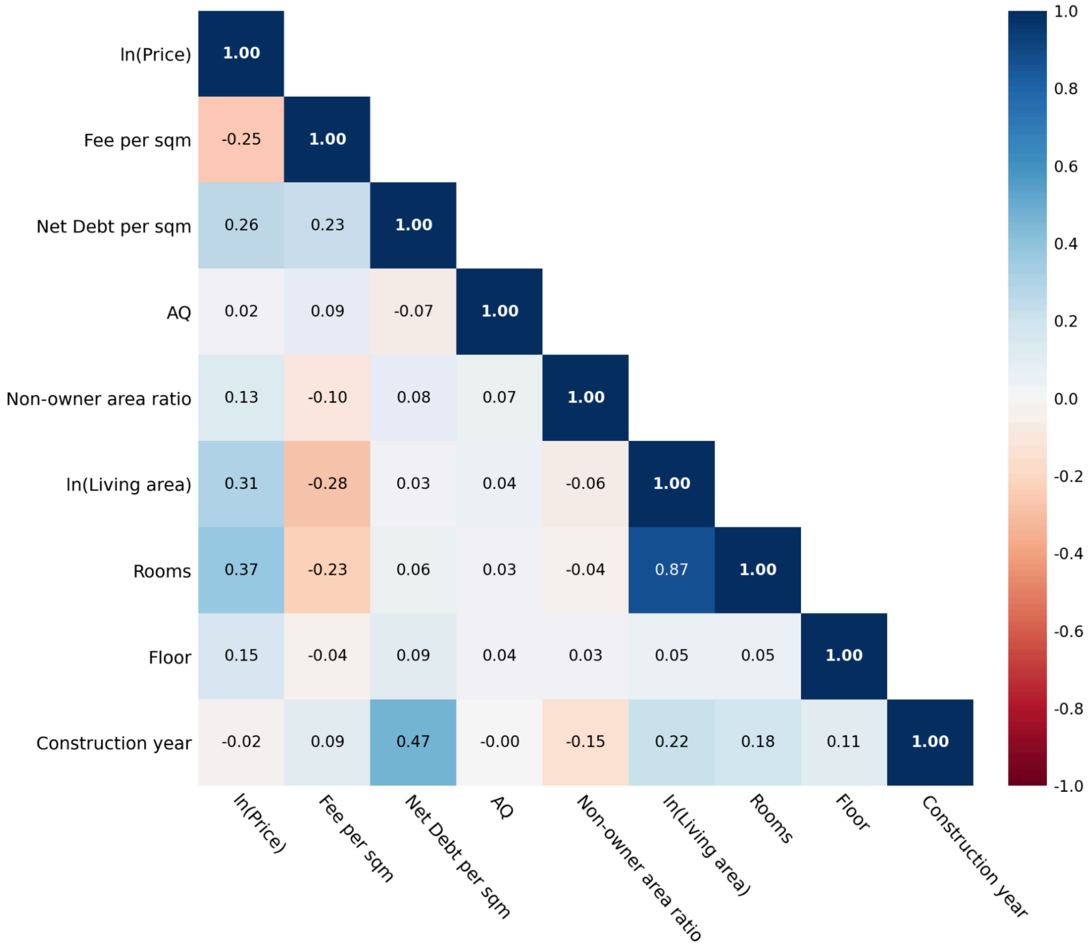
The H4 component decomposition provides no evidence of a standalone transparency premium for any individual disclosure dimension. AQ_A enters at −0.0028, AQ_B at $+7.5 \times 10^{-5}$, and AQ_C at −0.0010, with none reaching conventional levels of statistical significance. This is consistent with the weak main-effect result for the composite AQ measure.

The component-specific H5 interaction checks are all negative and statistically significant. The A-only, B-only, and C-only debt interactions are estimated at -1.7×10^{-6}, -1.9×10^{-6}, and -1.1×10^{-6}, respectively, each larger in absolute magnitude than the composite-AQ interaction in the baseline specification (-6.4×10^{-7}). This is consistent with the main H5 result and suggests that the moderation effect is not purely an artefact of aggregating the three AQ components.

A.10. Multicolinearity diagnostics

Figure A10.1 reports the Pearson correlation matrix for the main continuous regression variables. The matrix provides an initial descriptive check of pairwise correlations among the regressors. The strongest pairwise association is between log living area and rooms, which is expected given their close mechanical relationship. By contrast, the correlations among the main financial variables, fee, debt, and AQ, are modest. This suggests that the main financial regressors are not characterised by strong raw pairwise correlations.

Figure A.10.1. Pearson correlation matrix



Note: Pairwise Pearson correlations are computed on the common main-spec sample.

As a regression-specific diagnostic, variance inflation factors (VIFs) are reported for the main specifications to assess whether the estimated coefficients may be affected by problematic multicollinearity among the regressors (Kutner et al., 2018). Unlike the Pearson correlation matrix, which only captures pairwise correlations, VIFs assess whether each regressor is linearly predictable from the remaining regressors. The VIFs are computed on the fixed-effect-absorbed design matrix used in the corresponding regressions. Accordingly, the diagnostic reflects collinearity among the regressors after removing the variation explained by location and temporal fixed effects.

A VIF of 1 indicates no linear dependence on the remaining regressors, while higher values indicate increasing inflation in the variance of the coefficient estimate. Following common rules of thumb, VIF values above 10 may indicate potentially concerning multicollinearity (Kutner et al., 2018).

Table A9.1 shows that the main specifications generally do not suffer from problematic multicollinearity. The largest VIFs outside the interaction model are for log living area and rooms, at about 4.2-4.6. This is expected because apartment size and room count are mechanically related, but the values remain below conventional concern thresholds. The financial variables in the non-interaction specifications also have low VIFs where the fee and debt variables generally lie between about 1.2 and 2.1, indicating that the main fee and debt estimates are not driven by severe collinearity among the observed financial controls.

In the debt-by-AQ interaction specification, the VIF for net debt is 18.3 and the VIF for the interaction term is 19.1. Elevated VIFs are common in models with interaction terms because the interaction is mechanically related to its constituent variables. These values should therefore not be interpreted as evidence of broad multicollinearity across the model. Rather, the main debt coefficient and the interaction term should be interpreted jointly, with emphasis on the implied marginal effect of debt at different AQ levels, as is standard when interpreting interaction effects (Mize, 2019).

Table A.10.1. Summary of VIF results

Variable	**Fee** **Eq. (2)**	**Debt** **Eq. (3)**	**Fee + Debt** **Eq. (4)**	**Debt × Post2022** **Eq. (5)**	**AQ** **Eq. (6)**	**Debt × AQ** **Eq. (7)**
ln(Living area)	4.559	4.294	4.573	4.573	4.578	4.578
Rooms	4.241	4.236	4.247	4.247	4.248	4.248
Floor	1.019	1.017	1.019	1.019	1.019	1.019
New Building	1.213	1.223	1.224	1.252	1.232	1.248
Qualified association	1.001	1.006	1.007	1.007	1.007	1.007
Leasehold	1.021	1.027	1.036	1.036	1.036	1.036
NA ratio	1.020	1.028	1.044	1.044	1.049	1.049
Fee	1.202		1.330	1.334	1.337	1.337
Debt		1.213	1.343	2.161	1.346	18.322
Debt × Post2022				1.891		
AQ					1.026	2.826
Debt × AQ						19.106

Notes: VIFs are computed on the fixed-effect-absorbed design matrix used in the corresponding regression. Values above 10 are commonly interpreted as indicating potentially concerning multicollinearity. Empty cells indicate variables not included in the relevant specification.

As a complementary diagnostic, Figure A10.1 reports the Pearson correlation matrix for the main continuous regression variables. The matrix confirms that the strongest pairwise association is between log living area and rooms, which is expected given their close mechanical relationship. By contrast, the pairwise correlations among the main financial variables, fee, debt, and AQ, are modest. This supports the VIF results by showing that the main financial regressors are not characterised by strong raw pairwise correlations.

A.11. Disclosure of generative AI use

Three generative AI systems were used during the preparation of this thesis: Claude (Anthropic), ChatGPT (OpenAI), and Gemini (Google). We applied AI in two broader categories.

The first category concerned standard writing and learning support. The tools were used to reformulate and proofread passages, check grammar and clarity, explain unfamiliar statistical concepts, assist in identifying potentially relevant literature, translate text, and help write and debug Python code used in data processing and analysis. In all such cases, the underlying ideas, arguments, and interpretations are our own, and any literature suggested by the models was read and assessed independently before inclusion.

The second category covers the thesis's empirical data pipeline. As described in the method section, Claude operated within a multi-layer retrieval architecture as the agentic layer in the development phase. Generative AI was also used in the production run, for candidate site evaluation, evaluate report validity, extract information from reports and place AQ scores.

As mentioned previously in this paper, there were several risks with our use, but we ensured to address those we noticed. Generative models are known to produce plausible but inaccurate output, which poses a particular threat to empirical work. To mitigate this, the data produced by the evaluation, extraction, and qualitative-assessment steps was not accepted at face value. Section 3.2.2 documents the validation procedure applied to these outputs, including checks against the underlying source documents. Furthermore, as mentioned earlier, factual claims and citations suggested by the models in the writing process were verified against the original sources rather than incorporated directly. Code produced with AI assistance was reviewed and tested before use. Concerns regarding data protection were addressed by limiting prompts to publicly available information, with no personal or confidential data entered into any model.

For disclosure, The following prompts are illustrative of the kinds of instructions issued to the models during the project. They are reproduced here to give the reader a concrete sense of how the tools were used across the different task categories:

- *Language and proofreading:* "Please proofread the following paragraph for grammar, clarity, and academic tone. Do not alter the meaning or introduce new claims. Flag any sentences that are ambiguous rather than rewriting them silently."
- *Conceptual explanation:* "Explain the intuition behind using log for the dependent variable in a linear regression, and describe the assumptions under which it is appropriate. Point me to the standard references within the hedonic framework so I can read them directly."

- *Literature scoping:* "List control variables commonly used in the hedonic pricing framework, with peer-reviewed sources linked to each variable."
- *Coding assistance:* "Here is my Python function that parses a table from a PDF using pdfplumber. It returns empty rows for tables with merged header cells. Suggest why this might be happening and propose a fix."
- *URL selection within the retrieval agent:* "You are given a search query for a company's annual report and a list of candidate URLs with their titles and snippets. For each URL, judge whether it plausibly leads to the official annual report, an investor-relations page that may contain it, or is clearly unrelated. Return a short justification for each classification."

A.12. Swedish translation glossary

Much of this study centres around a Swedish setting. Accordingly, the table below defines the key Swedish terms used throughout the analysis.

Term	**Definition**
Cooperative apartment area	*Bostadsrättsyta* in Swedish. The usable floor area of apartments held with tenant-owner rights within the association.
Site leasehold	*Tomträtt* in Swedish. A leasehold arrangement where the association owns the building but leases the land from the municipality or another landowner.
TOA	*Bostadsrättsförening (BRF)* in Swedish. A tenant-owner association that owns and manages a residential property, while members hold the right to occupy individual apartments.
Qualifying housing cooperative	*Äkta förening* in Swedish. A tenant-owner association classified as a genuine housing association, mainly because its income primarily comes from members' residential use rather than commercial rental activity.