

Response to Reviewers

General Statement

We thank the Editor and the reviewers for their detailed and constructive feedback. We acknowledge that this revision focuses on resolving the remaining issues related to consistency, positioning, and the strengthening of the systematic literature review contribution.

All remaining points have been carefully addressed and are detailed below.

Response to Reviewer 2

Comment 1: Grossman–Stiglitz (1980) inconsistency

Comment. The rebuttal claims that the Grossman–Stiglitz (1980) reference and interpretation were corrected, but the manuscript still cites Grossman (1980) and the reference list remains incorrect.

Response. We thank the reviewer for pointing out this inconsistency. We have corrected the reference throughout the manuscript by replacing the SSRN entry, which only cited Grossman (1980), with the original published article by Grossman and Stiglitz (1980) in *The American Economic Review*. The in-text citation and the corresponding bibliography entry have been updated accordingly.

Comment 2: Residual inconsistency in Research Question 2

Comment. Although the manuscript reframes RQ2 as a conceptual synthesis question, remnants of the original formulation (“financial market dynamics”) remain in the screening statements and parts of the manuscript.

Response. We thank the reviewer for this careful reading.

To ensure full consistency with the revised formulation of RQ2, we reviewed the manuscript and corrected all remaining references that reflected the previous focus on financial market dynamics. In particular, screening statements that previously referred to “understanding” or “enhancing” market dynamics have been reformulated to align with the current objective of the review, namely the conceptualization and modeling of financial narratives.

In addition, we clarified the role of search terms related to market dynamics in the screening procedure. These terms were initially included to ensure broad coverage of the financial literature in which narratives are studied. We now explicitly state that their inclusion does not reflect the objective of the review, which is focused on how narratives are defined, represented, and modeled, rather than on their direct impact on markets.

Finally, the keywords and remaining minor inconsistencies have been updated to reflect this conceptual focus. These changes ensure that the research question, screening procedure, and manuscript content are now fully aligned.

Comment 3: Single-source limitation (Scopus)

Comment. The manuscript remains based solely on Scopus, which excludes SSRN, NBER, and arXiv.

Response. We thank the reviewer for raising this point. We agree that reliance on a single bibliographic source is an important limitation of the current implementation. As already stated in the manuscript, Scopus was retained because it offers a practical compromise between accessibility, metadata consistency, and scientific quality control. In the context of an algorithmic screening pipeline, this balance is important: abstracts, keywords, document types, and indexing fields are relatively standardized, which makes the semantic selection process more stable and reproducible than it would be on more heterogeneous repositories.

We would also like to clarify that this choice should not be interpreted as claiming that Scopus is exhaustive, nor that it is intrinsically superior to all other sources for every review design. Rather, it is the most accessible curated database for the specific purpose of this study, namely the construction of a reproducible and operational screening framework applied to peer-reviewed academic literature. By contrast, repositories such as SSRN, NBER, and arXiv provide broader and often more timely coverage, but they also introduce greater heterogeneity in document format, metadata quality, and review status. This would require an additional harmonization layer before semantic screening could be performed under comparable conditions.

As also discussed in the revised manuscript, a natural extension of the framework would be to rely on OpenAlex, which provides a unified API across multiple open and non-open scholarly repositories. This type of infrastructure would constitute a useful intermediary solution, as it broadens coverage while preserving a relatively structured metadata interface that remains compatible with automated retrieval and filtering.

Beyond this database-level extension, an even more ambitious improvement would be to move away from traditional keyword-constrained search endpoints and adopt vector-based retrieval directly at the corpus level. In such a setup, papers would be indexed through dense high-dimensional embedding vectors, allowing semantic search based on natural-language criteria rather than only lexical matching. The current re-ranking and clustering pipeline could then be applied on top of this retrieval layer. We view this as a promising direction because it would reduce dependence on database-specific query syntax and improve the identification of conceptually relevant studies expressed with different vocabularies.

For the present paper, however, we chose to keep the Scopus-based implementation because it provides a controlled and reproducible baseline for demonstrating the selection framework.

Comment 4: Subjectivity in inclusion criteria

Comment. The study relies on researcher-defined statements without validation or a ground-truth benchmark.

Response. We thank the reviewer for raising this point. We agree that the use of researcher-defined statements introduces a degree of subjectivity in the selection process, and we acknowledge that this is a structural aspect of any semantic retrieval framework that does not rely on a predefined ground-truth benchmark.

That said, several design choices were implemented to mitigate this subjectivity and reduce dependence on arbitrary wording. First, the embedding model used (`text-embedding-3-small`) is a large-scale, general-purpose model trained on a broad corpus, which allows it to capture semantic similarity beyond surface-level phrasing. In practice, this enables comparisons at the level of underlying concepts rather than exact keyword overlap, which is precisely the limitation of traditional

database queries.

Second, we introduced an explicit rephrasing step using a large language model. Each statement is automatically reformulated into multiple semantically equivalent variants, which are then embedded and aggregated. This reduces sensitivity to specific wording choices and limits the risk that inclusion decisions are driven by particular lexical formulations rather than conceptual alignment.

Third, the six statements used in the screening process were intentionally designed to be broad and complementary. Rather than targeting narrow definitions, they cover different dimensions of the research space (context, methods, data, and research questions). Their role is therefore closer to structured semantic filters than to rigid inclusion rules. Compared to standard keyword-based queries, this approach provides wider coverage while maintaining a more precise alignment with the research objective.

More broadly, this limitation is partly inherited from the constraints of current literature databases, where retrieval is still anchored to query-based systems. As noted in our response to the previous comment, a more robust setup would consist of indexing the entire corpus as a vector database and performing retrieval directly in the embedding space, followed by the re-ranking procedure proposed in this paper. Such a framework would allow systematic sensitivity analyses with respect to phrasing and statement design at scale, which is not feasible under standard query-based access.

Finally, we agree that a formal validation protocol would further strengthen the study. For instance, one could construct a benchmark dataset with manually labeled relevance or compare retrieval outcomes across alternative statement formulations. However, implementing such a validation framework would require a substantial annotation effort and would constitute a research contribution on its own, particularly in studying how semantic phrasing affects vector-based retrieval in scientific corpora. While this is beyond the scope of the present paper, we explicitly acknowledge it as a meaningful direction for future work.

Comment 5: Exclusion of medium-relevance papers

Comment. Potentially relevant studies are excluded by design, which may introduce bias.

Response. We appreciate the reviewer’s observation regarding the potential bias introduced by excluding medium-relevance papers. This is a valid concern, particularly in research where the boundary between relevance levels is not clearly defined.

In our use case, this design choice was not applied blindly. As part of the refinement process, we conducted a manual inspection of papers located near the decision boundary between high- and medium-relevance clusters. This inspection confirmed that a large share of these papers did not directly address the research questions, but rather mentioned related concepts in a peripheral or non-substantive way. The exclusion of these papers was therefore intended to improve the signal-to-noise ratio of the final corpus.

More importantly, this aspect should be understood as a configurable component of the methodology rather than a fixed constraint. The framework is designed to support different levels of granularity depending on the objective of the review. In particular, users may opt for: (i) a binary classification (high vs. low relevance), which maximizes coverage, or (ii) a three-tier classification (high, medium, low), which allows for stricter filtering and more focused synthesis.

The appropriate choice depends on several factors, including the breadth of the research field, the maturity and density of the literature, and the specific objectives of the review. In well-established and highly populated domains, introducing an intermediate class and excluding medium-relevance papers can help maintain interpretability and reduce dilution effects. Conversely, in emerging or sparsely populated fields, retaining these papers may be preferable to avoid missing relevant

contributions.

We have clarified this flexibility in the manuscript to emphasize that the exclusion of medium-relevance papers is not an inherent limitation of the approach, but rather a methodological option that can be adapted to the context of the study.

Response to Reviewer 5

Major Comment: Strengthening concept-centric synthesis and gap identification

Comment. The manuscript under-develops the synthesis and gap-identification component. A concept-centric framework (e.g., TCM, concept matrix) should be introduced to structure the review and explicitly derive research gaps.

Response. We thank the reviewer for this suggestion, which directly improved the structure and clarity of the manuscript.

To address this point, we introduced a dedicated subsection titled *Concept-centric synthesis and gap identification*, where the literature is now structured using the Theory–Context–Method (TCM) framework proposed by Paul and Criado (2020). We added, after the descriptive review of individual papers, the synthesis organized along three explicit dimensions: (i) the theoretical conceptualization of narratives, (ii) the empirical context in which they are studied, and (iii) the methodological approaches used for their extraction and integration into financial models. We constructed the corresponding concept matrix that maps all selected papers across these three dimensions. This matrix provides a compact and systematic representation of the literature, making it possible to compare studies along common axes.

Building on this structure, the identification of research gaps is now derived directly from the absence of specific combinations within the matrix. In particular, the revised manuscript highlights the lack of studies combining structured narrative representations, advanced semantic methods, and integrated financial modeling, as well as the limited transfer of recent NLP advances to financial applications and the narrow empirical scope of existing work.

This revision strengthens the synthesis by improving the descriptive aggregation of studies with a concept-centric analysis, and ensures that the research gaps are derived in a transparent and systematic manner.

Minor Comment 1: Role of AI in the synthesis stage

Comment. The manuscript does not clearly explain how AI supports data extraction and reduces manual effort in the synthesis stage.

Response. We thank the reviewer for raising this point.

The present work is intentionally focused on the search, filtering, and reranking stages of a systematic literature review. The proposed framework is designed as an AI-enhanced selection pipeline, and its contribution lies in improving how relevant papers are identified and prioritized.

In this scope, AI is not used in the synthesis stage. Extending the framework to support automated or semi-automated synthesis would require a different class of methods, such as structured information extraction, topic modeling, clustering, or large language models applied to full-text analysis. These approaches introduce additional methodological challenges. Topic modeling typically provides coarse representations that are not sufficient for rigorous concept-level synthesis, while large language models remain sensitive to prompting and can introduce hallucination or confirmation bias when asked to identify research gaps.

Addressing these issues would require a dedicated research effort, including the design of validation protocols and robustness checks, which is beyond the scope of the current paper. The objective here is therefore deliberately restricted to demonstrating how embedding-based retrieval and reranking can improve the upstream stages of the review process.

That said, the results suggest that combining vector-based retrieval with adapted reranking techniques provides a scalable foundation that could, in future work, be extended toward more advanced synthesis pipelines, for instance within vector database environments where similarity scoring and clustering can be applied at scale.

Minor Comment 2: Editorial issues

Comment. Typographical issues remain (e.g., “Section ??”, “words presences”).

Response. We thank the reviewer for pointing out these issues. The incorrect section reference previously displayed as “Section ??” has been corrected, and the phrasing “words presences” has been revised for clarity and correctness. In addition, we conducted a careful review of the manuscript to improve language quality, consistency, and overall academic style.

Additional Consistency Corrections

Consistency of Research Questions and Methodology

We performed a full consistency check across the manuscript to ensure alignment between research questions, methodology, screening criteria, and discussion. In addition to the targeted corrections described above, we performed a full consistency check across the manuscript. The year filter is now uniformly expressed as “2011 or later” in both the selection criteria table and the results section. The sixth screening statement has been updated to reflect the revised RQ2, replacing the previous reference to “financial market dynamics understanding” with the current formulation focused on how narratives are defined and applied. A clarifying sentence has been added in Section 2.1 to make explicit that search terms related to market dynamics were included solely to ensure broad coverage of the financial literature, and do not reflect the conceptual objective of the review.

Conclusion

We believe that this revision addresses all remaining concerns raised by the reviewers. The manuscript has been corrected for consistency issues and strengthened in its synthesis and gap-identification components.

We thank the reviewers again for their valuable feedback.