Author Response to Round 2 Comments

Manuscript: NarrativeBreak: Integrating Structural Break Detection with Multi-Source NLP Signals for Dynamic Portfolio Optimization **Authors:** Joerg Osterrieder **Response Date:** Day 135 (Simulated)

We thank the reviewers for their continued constructive feedback. All comments have been addressed in this revision.

Response to Reviewer 1

1. HMM Regularization

Comment: Request for explicit alpha values and sensitivity analysis.

Response: [ADDRESSED] We have added the following to Section 3.4:

"We use Dirichlet priors with alpha = 1.0 (uniform prior) for each row of the transition matrix. With our 2,500+ day training sample, this regularization has minimal effect on posterior estimates but prevents numerical degeneracy in edge cases. Sensitivity analysis with alpha in {0.1, 0.5, 1.0, 2.0, 5.0} showed no qualitative change in results (transition probabilities varied by < 0.02)."

2. Appendix Organization

Comment: Suggest more descriptive appendix titles.

Response: [ADDRESSED] Appendices renamed: - Appendix A: "Synthetic Data Calibration and Validation" - Appendix B: "Real-Data Validation: March 2020 COVID-19 Period"

3. Table A1 Presentation

Comment: Add sample size column and clarify p-value interpretation.

Response: [ADDRESSED] Table A1 now includes: - Sample size column for each statistic - Footnote: "High p-values indicate failure to reject the null hypothesis that synthetic and target distributions are identical (desirable outcome)."

4. Lead Time Confidence Interval

Comment: Mention wide CI in abstract/introduction.

Response: [ADDRESSED] The abstract now reads: "...detecting narrative regime changes 5.7 days (95% CI: 1.2–10.2) before price-based methods."

Response to Reviewer 2

1. Abstract Framing

Comment: Abstract should reflect Bonferroni-adjusted significance.

Response: [ADDRESSED] The abstract has been revised to use cautious language:

"...achieves a Sharpe ratio improvement of 83% over equal-weight portfolios, with 26% reduction in maximum drawdown. While these improvements are economically meaningful, statistical significance is marginal after multiple testing correction."

2. Conclusion Section

Comment: Include note about marginal significance.

Response: [ADDRESSED] Section 7 now includes:

"We note that while our results show economically meaningful improvements, statistical significance is sensitive to multiple testing corrections (Bonferroni-adjusted $p = 0.17$). We therefore characterize our findings as 'suggestive evidence' warranting further validation on out-of-sample real data."

3. GPT-4 Discussion

Comment: Expand LLM discussion with version, cost trajectory, and alternatives.

Response: [ADDRESSED] Section 5.4 now includes:

"We tested GPT-4 (gpt-4-0613) on a 500-sample subset, achieving ~70% accuracy at approximately $0.02 per sample (vs. <$0.001 for FinBERT). While inference costs have decreased substantially since our testing (GPT-4o offers comparable performance at ~10x lower cost), open-source alternatives such as Llama-3 and Mistral may offer better cost-accuracy tradeoffs for production deployment. We leave systematic LLM comparison to future work."

4. Real-Data Case Study Limitations

Comment: State limitations more explicitly.

Response: [ADDRESSED] Appendix B now includes:

"Limitations: (1) GDELT data quality varies by source; we filtered to major outlets (Reuters, Bloomberg, WSJ) for consistency; (2) Entity recognition was performed manually for this case study, which is not scalable; (3) The 2-month window is too short for HMM regime detection to converge reliably—we used simplified threshold-based detection as a proxy. These limitations mean the case study should be viewed as qualitative validation only."

5. Minor Typos

Response: [ADDRESSED] All typos corrected: - "behaviour" -> "behavior" (Page 7, line 23) - Added period after "days" in Table 5 - Reference [24] abbreviation standardized

---

Summary of Round 2 Changes

1. Abstract revised with confidence interval and cautious significance language
2. Conclusion updated with statistical caveats
3. Section 3.4 expanded with HMM regularization details
4. Section 5.4 expanded with LLM alternatives discussion

5. Appendices renamed with descriptive titles
6. Table A1 enhanced with sample sizes and interpretation footnote
7. Appendix B limitations stated explicitly
8. Minor typos corrected throughout

We believe the manuscript is now ready for publication and thank both reviewers for their constructive engagement throughout the review process.

Joerg Osterrieder Bern University of Applied Sciences University of Twente

---

Note: This is a simulated response for demonstration purposes.