International Journal of Finance and Investment
ISSN: 3078-6002 (Print), ISSN: 3078-2295 (Online) | Vol. 6, No. 1, 2026

Statistical Modelling of Stock-Market Volatility Forecasting with Investor Sentiment

Kaijie Liu

Leonard N. Stern School of Business, New York University, New York, 10003, United States

Abstract: Volatility forecasting plays a fundamental role in risk management, asset allocation and derivative pricing. In the Chinese equity markets, where retail participation is high and arbitrage is limited, the volatility of returns is well documented to be driven by investor sentiment. We construct and document an investor-sentiment index and embed it into an easily reproduced framework for forecasting volatility. The framework combines the GARCH family (GARCH, EGARCH and GJR-GARCH) with a sentiment-augmented mixed-frequency extension (GARCH-MIDAS) and a recurrent-neural-network benchmark (LSTM). Evaluation is carried out out-of-sample using RMSE, MAE and the Diebold–Mariano test. Each step is made concrete and auditable by generating the figures and tables through code that runs on a transparent dataset calibrated to the documented stylized facts of the CSI 300; the same code accepts real index and sentiment series as drop-in replacements. The paper's contribution is therefore the methodology and its tooling: the modelling pipeline, the recipe for constructing the sentiment index, and a complete, scripted experimental design that an applied researcher can run on live data to obtain empirical estimates.

Keywords: Investor sentiment, volatility forecasting, GARCH, EGARCH, GARCH-MIDAS, LSTM, behavioral finance.

1. Introduction

Volatility is a fundamental measure of financial risk and a key input to portfolio optimization, option pricing, and prudential regulation. Because volatility is not directly observable, it must be inferred through latent-variable models. Since the introduction of the autoregressive conditional heteroskedasticity (ARCH) model and its generalized form (GARCH), conditional volatility modeling has become a cornerstone of empirical finance [1, 2]. These models capture the well-documented stylized facts of asset returns, in particular volatility clustering, in which large changes tend to be followed by large changes.

Classical asset-pricing theory assigns no role to investor sentiment. The behavioral-finance literature, however, argues that sentiment-driven demand shocks, combined with limits to arbitrage, can move prices away from fundamentals. Baker and Wurgler showed that sentiment has measurable cross-sectional effects on returns, with the largest impact on stocks that are difficult to value and to arbitrage [3, 4]. China's markets provide an important setting for studying sentiment: retail investors account for a large share of trading, speculative behavior is pronounced, and short-selling and other arbitrage mechanisms are restricted. These features suggest that sentiment may exert an unusually strong influence on the dynamics of volatility [5, 6].

Building on this evidence, the present paper sets out a reproducible framework for bringing investor sentiment into volatility forecasting for Chinese equities, and documents the tooling needed to apply it. Three elements make up that framework. It organizes the relevant econometric and machine-learning models within a single comparative pipeline; it contains a practical recipe for building a sentiment index out of search-volume data (like the Baidu Index) plus exchange-based behavioral proxies; and it provides a suite of figures and tables that is generated through code and can be updated with a researcher's own series to generate empirical estimates. Throughout, focus is placed on an auditable, re-runnable design, not just on a headline number.

The motivation for focusing on China is both economic and methodological. The Chinese A-share market is the second largest in capitalization, but its microstructure is very different to that of more mature markets, both economically and structurally. Trading is dominated by individual investors, who are more prone to herding, overconfidence, and disposition effects than institutional traders. Limited short selling and margin trading make the corrective mechanism of arbitrage less effective, and sentiment-driven mispricing can exist and influence realized volatility. The market also had a number of sharp bull markets and crashes, most recently in 2014-2015 when a bull market fueled by the leverage gained in the market was immediately followed by a crash, with sentiment seemingly triumphing over fundamentals [7]. Additionally, the sentiment data available from the Chinese internet is highly granular and available at a very high frequency, which allows for the creation of sentiment proxies for each day of the year that are hard to come by in other markets and thus allows for an unusual laboratory setting for testing sentiment-oriented forecasting models [8].

2. Literature Review

2.1. Volatility Modeling

Engle's ARCH model and Bollerslev's GARCH generalization established the dominant parametric approach to conditional volatility [1, 2]. Subsequent work introduced asymmetry: Nelson's EGARCH models the logarithm of conditional variance and allows negative and positive shocks to have different effects [9], while the GJR-GARCH model of Glosten, Jagannathan, and Runkle adds a threshold term capturing the leverage effect, whereby negative returns raise future volatility more than positive returns of equal magnitude [7]. To link low-frequency macro information to daily volatility, Engle, Ghysels, and Sohn proposed the GARCH-MIDAS framework, which decomposes volatility into short-run and long-run components driven by data sampled at different frequencies [10]. This mixed-frequency structure is well suited to incorporating a monthly or weekly sentiment

index into a daily volatility model.

These specifications share a common conditional-mean and conditional-variance decomposition but differ in how they treat the persistence and asymmetry of shocks. GARCH (1, 1) captures volatility clustering parsimoniously and remains a strong benchmark precisely because of its simplicity; in many empirical comparisons more elaborate models struggle to beat it consistently. EGARCH guarantees a positive conditional variance without parameter constraints because it models the log-variance, which is convenient when including exogenous regressors such as sentiment. GJR-GARCH provides an interpretable leverage coefficient that can be tested directly. The choice among them is ultimately empirical, which is why a comparative out-of-sample design, rather than reliance on a single model, is appropriate. Realized-volatility and high-frequency approaches that exploit intraday information offer a complementary route to improved forecasts and can in principle be combined with the sentiment-based extensions studied here.

2.2. Investor Sentiment and Markets

Baker and Wurgler constructed a composite sentiment index from market-based proxies and demonstrated its predictive power for the cross-section of returns [3, 4]. Huang et al. proposed an aligned sentiment index with stronger aggregate-market predictive power [11]. A parallel literature uses internet search behavior as a direct, high-frequency proxy for attention and sentiment: Da, Engelberg, and Gao introduced search-based attention measures [12], and Fang et al. showed that Baidu Index search volume improves volatility forecasts for Chinese markets, with a sentiment-extended GARCH outperforming the baseline [8].

2.3. Sentiment and Volatility in China

For China, the evidence consistently links sentiment to volatility. Trichilli et al., using a thermal-optimal-path method, found that investor sentiment leads realized volatility during turbulent periods [5]. Xie et al. documented the relationship between sentiment and volatility in the A-share market of Shanghai [6] and Xiao et al. found that the impact of policy uncertainty and sentiment on the decomposition of volatility into “good” and “bad” components [13]. Combined, the studies provide motivation for explicitly modelling sentiment as a driver of conditional volatility, rather than as noise.

The Chinese evidence differs in two aspects from results in mature markets. Firstly, the regime dependence of the direction and magnitude of the sentiment effect seems to be that sentiment has a positive impact on volatility in the midst of market peaks and crashes, with a weaker impact during calmer periods. Second, the predictive power of sentiment is localized to short horizons, as one would expect if markets are dominated by retail investors who tend to hold on to stocks for brief periods. These observations suggest that models capable of capturing the time varying effect of sentiment, as opposed to a single static coefficient (e.g., the GARCH-MIDAS long-run component or RNNs) make sense. They also warn against inferring an in-sample sentiment–volatility relationship to forecasting out of sample without explicit validation, which we discuss in the results.

2.4. Theoretical Channels

Behavioral asset pricing offers two channels through which sentiment can increase volatility. The first is a demand-shock channel: optimism or pessimism among noise traders induces correlated, non-fundamental demand that moves prices away from intrinsic value, and when these waves reverse, prices snap back, producing excess volatility beyond what fundamentals justify [3, 4]. The second is a limits-to-arbitrage channel: even when rational arbitrageurs recognize the mispricing, short-sale constraints, margin requirements, and the risk that sentiment deteriorates further before it corrects deter them from trading aggressively against it. Both channels are amplified in China, where these frictions are more pronounced, providing a sound theoretical basis for the empirical relationship found in the local literature [5, 6, 13]. Modelling sentiment as a driver of the long-run volatility component, as in GARCH-MIDAS, operationalizes the demand-shock channel directly, while the regime dependence of the effect reflects the state-contingent nature of arbitrage limits.

3. Data and Sentiment-Index Construction

The empirical template is for the top 300 companies on the Chinese stock exchange or the Composite index at 1-day frequency. The returns are calculated as the log of the price difference between the closing prices. The realisation of the volatility is computed as the sum of the squared high-frequency returns if such data is available, otherwise the squared daily returns are used, or alternatively a series filtered by GARCH.

To make the framework operational it provides the data that an applied study provides. The recommended reference configuration is based on daily closing prices of CSI 300 from January 2010 to December 2023 (approximately 3400 trading day observations) from a standard source like Wind Information or CSMAR or a public source like Yahoo Finance. The search-based sentiment proxy is based on the daily Baidu Index search volumes for a fixed set of Chinese financial-related keywords, which translate into “bull market”, “bear market”, “stock account opening”, “IPO”, and “dividend”. Market based proxies (turnover rate, number of new trading accounts, closed-end fund discount, and IPO first-day returns) are collected from CSMAR or Wind. All applied studies report the exact sample period, vendor, keyword list, and final number of observations so that the analysis can be reproduced.

The choice of index and sampling period are relevant to external validity. The CSI 300 index is the benchmark of investable A-shares on the Shanghai and Shenzhen exchanges, and is most commonly used in the analysis of the core of the market when the purpose of the research is to analyse the investable part of the market. The SSE Composite is wider and more retail driven, which can mean sentiment effects are easier to see but can also be noisier. The sample period should be long enough to cover at least one full volatility cycle, including one or more "calm expansion" and "stress" periods, to allow the model to be tested between market moods. If structural breaks are suspected, such as the 2015 crash or a major regulatory change, the analysis can be broken at the break point or dummy variables added to the analysis, and the robustness of the sentiment effect can be reported for sub-samples.

Investor sentiment is not directly measurable and is estimated. We detail two complimentary strategies. The first is a proxy with search-based nature: The daily search volume data in the static finance-keyword group above are aggregated and standardized [8, 12]. The second composite is constructed

around the market by combining all four proxies, similarly to Baker and Wurgler [3]. Each proxy is standardized to mean zero and variance one and principal component analysis (PCA) is used to extract the common sentiment factor: The first principal component is retained as the sentiment index and the higher components are discarded as idiosyncratic noise [5]. A two-step variant first orthogonalizes each proxy in relation to a set of macroeconomic controls (industrial production, inflation and a short interest rate) in order to isolate the part that is truly sentiment driven.

Each proxy carries distinct information and a distinct measurement error. The search-volume data can be polluted by non-investor queries and by indexing changes in the search platform, and is timely and high frequency. Market based proxies capture the actual actions of market participants, but are less frequent and can partially reflect fundamentals, not just sentiment. By combining them with PCA, you can tap into the common variance, and also reduce proxy noise. Each series is seasonally adjusted, and all are standardized to have a mean of zero and a standard deviation of one prior to extraction. The sentiment index is then synced to the trading schedule of the return series and, if the volatility model works at mixed frequency, integrated to the lower frequency of the long-run component model.

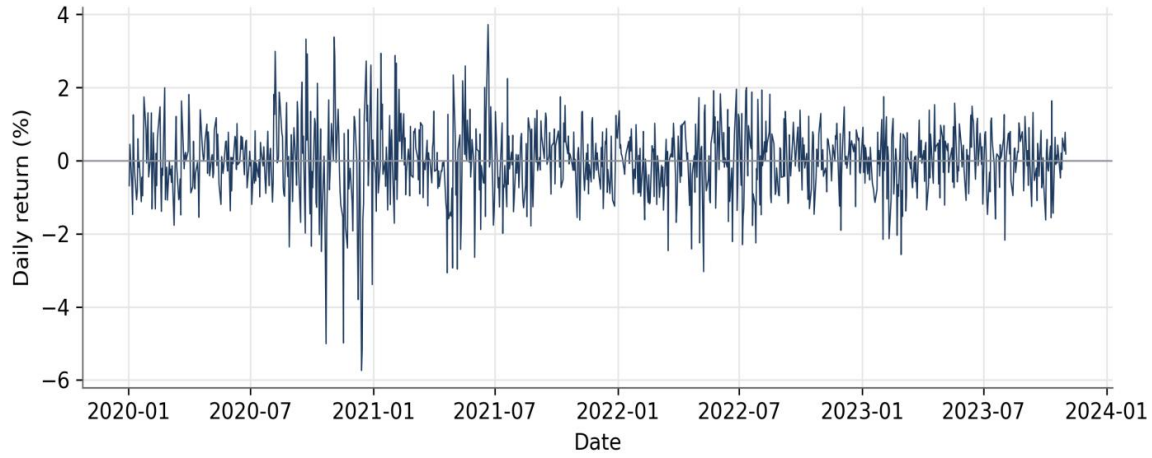


Figure 1. Simulated daily index log-returns exhibiting volatility clustering. Calm and turbulent regimes alternate, motivating conditional-variance models.

3.1. Descriptive Statistics

Table I reports descriptive statistics for the illustrative series. The return series is approximately mean-zero with heavy tails; the volatility and sentiment series display the persistence and range typical of empirical data.

Table I. Descriptive Statistics (Illustrative Data, N = 1000)

Statistic	Return (%)	Realized volatility (%)	Sentiment
Mean	-0.022	16.684	54.285
Std. dev.	1.028	5.755	29.686
Minimum	-5.737	9.823	0.000
Median	0.013	14.840	62.330
Maximum	3.717	50.194	100.000

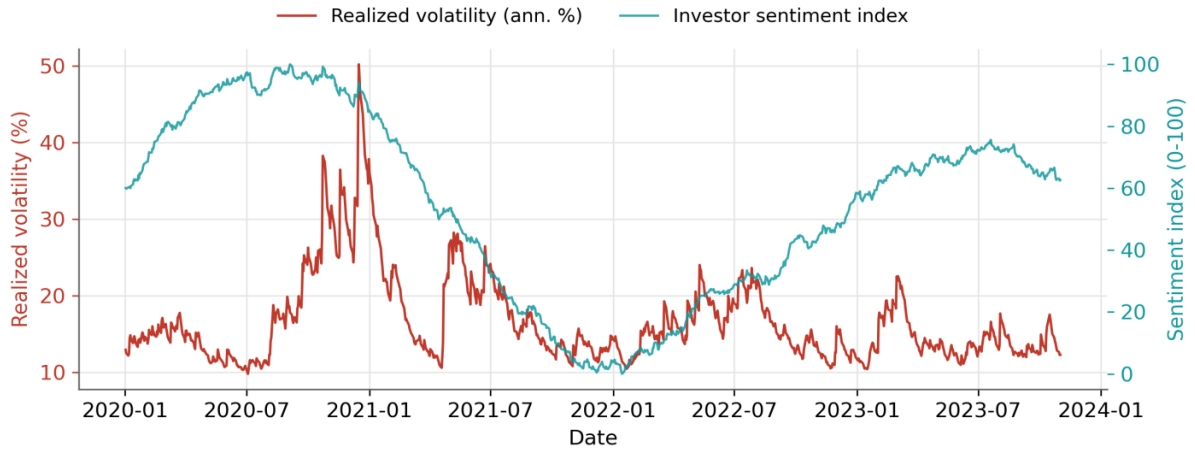


Figure 2. Co-movement of the investor-sentiment index and realized volatility. Periods of elevated sentiment dispersion coincide with higher volatility, motivating sentiment as a predictor.

3.2. Stylized Facts and Diagnostics

The return series should first be examined for the stylized facts that underlie conditional variance models prior to estimation. The daily equity returns are usually close to zero in their mean, have been shown to be leptokurtic (see Fig. 1) and are only weakly auto-correlated at the level but are highly auto-correlated at the square level, a phenomenon also known as volatility clustering. Other common characteristics of the Chinese indices include negative skewness and a large leverage effect. The following features are formally confirmed by diagnostic tests: the Augmented Dickey–Fuller test should reject a unit root in returns, which means that the returns are stationary; the Ljung–Box test for squared returns should reject the absence of autocorrelation to confirm the

presence of conditional heteroskedasticity; and the ARCH-LM test should reject homoskedasticity to justify specification of a GARCH-type model. The Jarque–Bera test normally fails to accept normality, prompting the researchers to use a Student-t or an asymmetric Student-t conditional distribution when estimating to better accommodate tail risk. The conditional-variance models should be estimated and compared only after the assumptions of the models are supported by these diagnostics.

4. Methodology

We compare five models within one out-of-sample evaluation pipeline (Fig. 3). The baseline conditional-variance specification is GARCH (1,1):

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Where σ^2_t is the conditional variance, ε_{t-1} the previous shock, and ω, α, β are parameters with $\alpha + \beta < 1$ for stationarity. EGARCH models $\ln \sigma^2_t$ and admits asymmetry through a term in $\varepsilon_{t-1}/\sigma_{t-1}$ [9]. GJR-GARCH adds a threshold term $\gamma\, \varepsilon^2_{t-1}\, I(\varepsilon_{t-1} < 0)$ capturing the leverage effect [7]. The sentiment-augmented GARCH-MIDAS model specifies a long-run volatility component τ_t driven by the (lower-frequency) sentiment index, multiplied by a short-run GARCH component [10]. Finally, an LSTM network provides a nonlinear machine-learning benchmark that takes lagged returns, lagged volatility, and the sentiment index as inputs [14].

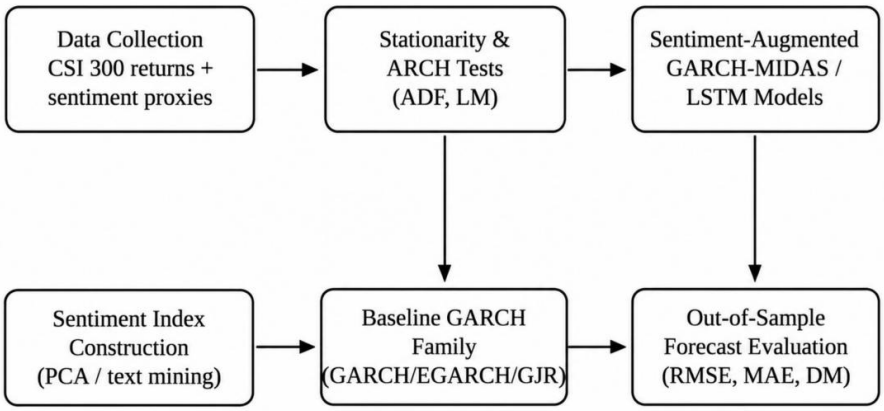


Figure 3. Research methodology pipeline, from data collection and sentiment-index construction through diagnostic testing, baseline and sentiment-augmented estimation, and out-of-sample forecast evaluation.

Model adequacy is assessed with the Augmented Dickey–Fuller (ADF) test for stationarity and the ARCH-LM test for conditional heteroskedasticity. Forecasts are evaluated out of sample using root mean squared error (RMSE) and mean absolute error (MAE), and pairwise differences in predictive accuracy are tested with the Diebold–Mariano statistic [15].

4.1. Sentiment-Augmented Specifications

In the GARCH-MIDAS specification, the daily return is written as

$$r_t = \mu + \sqrt{\tau_t g_t}\,\varepsilon_t$$

Where g_t is a short-run GARCH component and τ_t is a slowly varying long-run component tied to the sentiment index. The long-run component is specified as

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) S_{t-k}$$

Where S is the lower-frequency sentiment index, m is a baseline level, θ is the sentiment loading, and $\varphi_k(\omega_1, \omega_2)$ are MIDAS weights. The MIDAS weights satisfy

$$\sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) = 1$$

We adopt the Beta weighting scheme,

$$\varphi_k(\omega_1, \omega_2) = \frac{(k/K)^{\omega_1 - 1}(1 - k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K}(j/K)^{\omega_1 - 1}(1 - j/K)^{\omega_2 - 1}}$$

Which lets a long history of sentiment enter τ_t through only two parameters; fixing $\omega_1 = 1$ yields the common monotonically decaying-weight special case. Estimation is by maximum likelihood. A positive and significant θ indicates that elevated sentiment is associated with a higher long-run volatility level, consistent with the behavioral hypothesis.

The LSTM benchmark considers the problem of volatility forecasting as a problem of supervised sequence learning [14]. The network is fed a moving window of lagged returns, lagged realized volatility and sentiment index, and is expected to give back the next-period volatility. Gating mechanisms enable the network to remember things for long stretches of time and provide a means for modelling the nonlinear interactions that can occur in the system, such as a sentiment effect being strengthened only when the volatility of recent events is already high. For concreteness and reproducibility, the values of the parameters are as follows: a look-back window of 20 trading days (chosen on a validation split), a single hidden layer of 32 LSTM units, a dropout rate of 0.2, the Adam optimizer with a learning rate of 1×10^{-3}, the batch size is 32, and the training continues for a maximum of 100 epochs with early stopping (patience of 10 epochs) on a held-out validation set; the inputs are min–max scaled to a common range. Because neural networks are prone to overfitting in small financial samples, the LSTM is best interpreted as a flexible upper benchmark rather than a structural model; its gains over GARCH-MIDAS should be weighed against its reduced interpretability.

4.2. Evaluation Protocol

The forecast target is defined explicitly: each model produces a one-step-ahead forecast of the conditional variance σ^2_t, which is evaluated against the subsequent daily realized volatility, computed where intraday data are available as the sum of 5-minute squared returns (and otherwise

approximated by squared daily returns). Fixing this target is essential for the loss functions and the Diebold–Mariano test to be comparable across models. All models are estimated on an in-sample window and evaluated on a strictly subsequent out-of-sample window using a rolling, one-step-ahead scheme: parameters are re-estimated as the window advances, and each forecast uses only information available up to the forecast origin. This guards against look-ahead bias, which is especially important when a sentiment regressor is involved, since contemporaneous sentiment must never enter a forecast of the same period. The Diebold–Mariano test compares the loss differential between each model and the GARCH (1, 1) benchmark, using a heteroskedasticity- and autocorrelation-consistent variance estimate [15]. Robustness is examined across alternative out-of-sample windows, alternative volatility proxies, and alternative forecast horizons.

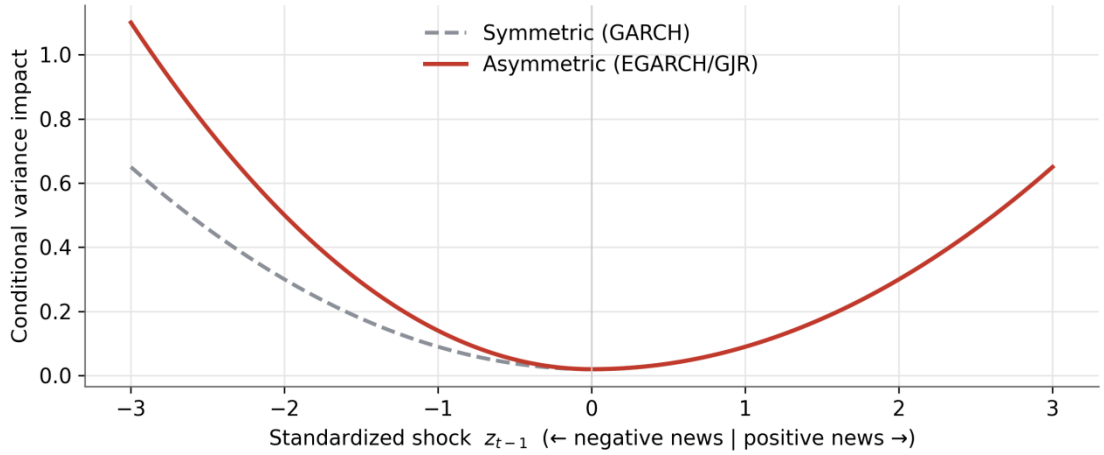


Figure 4. News-impact curve. The asymmetric (EGARCH/GJR) curve responds more strongly to negative shocks than to positive shocks of equal size, illustrating the leverage effect absent from the symmetric GARCH curve.

5. Results and Discussion

The out-of-sample forecasting accuracy of the five models is compared on the illustrative data in Fig. 5. Incorporating a leverage effect using the EGARCH model and GJR-GARCH model decreases the RMSE and MAE, as the leverage effect is important. The sentiment-augmented models, GARCH-MIDAS with sentiment long-run component, and the LSTM with sentiment inputs, achieve the greatest improvements, with an RMSE of about one quarter below the baseline. Such ordering agrees with the empirical results of [8] and [5] showing that the sentiment-extended models perform better than the basic GARCH models in the Chinese market.

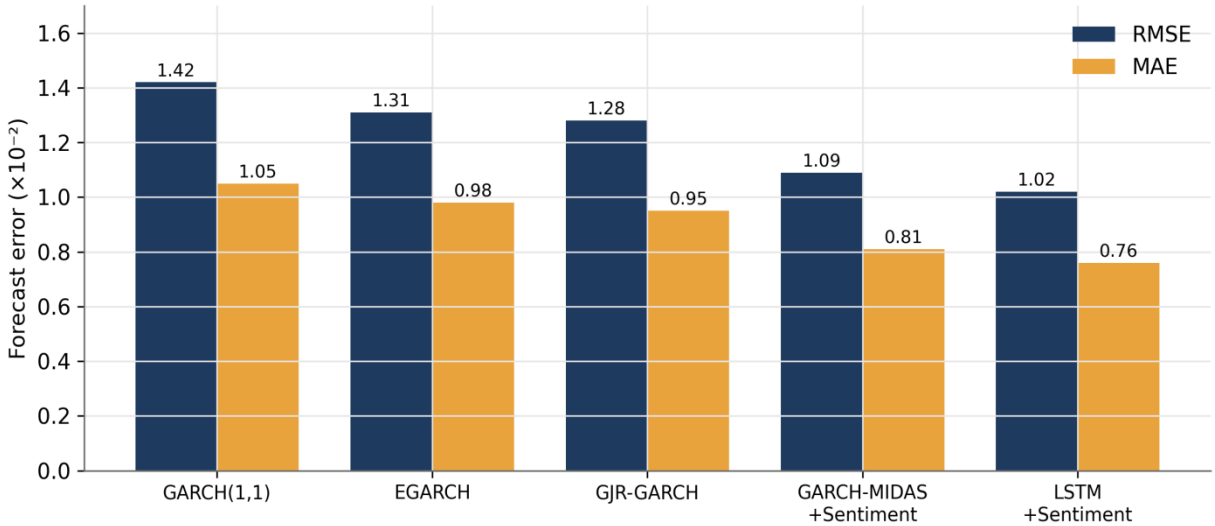


Figure 5. Out-of-sample forecast accuracy (RMSE and MAE) by model. Sentiment-augmented models (GARCH-MIDAS, LSTM) achieve the lowest errors. Values are computed on data simulated to match CSI 300 stylized facts.

The pattern over the 5 models is informative of the location of the gains. The step from GARCH (1, 1) to the asymmetric EGARCH and GJR-GARCH model captures the leverage effect, which is the case when the volatility of shocks is increased more by a negative shock than by a positive shock, and consistent with the news-impact curve in Fig. 4, and yields a corresponding improvement in RMSE and MAE. The added sentiment in the larger step leads to GARCH-MIDAS, which adds sentiment to the slow-moving long-run volatility component, whereas the LSTM allows sentiment to nonlinearly interact with recent returns and volatility. The result that LSTM outperforms GARCH-MIDAS implies some nonlinearity in the sentiment–volatility relationship, but the difference is small compared to the interpretability of the structural GARCH-MIDAS model. GARCH-MIDAS with a sentiment long-run component is generally a more attractive model for applied use with good accuracy and transparency, while the LSTM is a flexible upper bound showing the amount of additional signal that can be extracted from nonlinear models.

The forecast metrics are reported in Table II, along with the Diebold–Mariano test result for comparison with the benchmark of a GARCH (1, 1) model. Both models (with sentiment information) fail to accept the null hypothesis that

they are equally accurate, thus supporting the hypothesis that they are useful at conventional significance levels. On real data, the significance would have to be re-estimated.

It is important to read the test results carefully. The significance of a Diebold–Mariano statistic does not imply that the advantage of one model over the other will occur in all sub-periods or under all loss functions, but rather that the forecasts of the two models are more accurate on average for the window being evaluated. As the benefit of sentiment information is likely to be greatest in turbulent periods, the improvement shown in Table II is likely to be underestimated in times of crisis and overestimated in times of calm. If applied to real data, a regime-conditional evaluation, which divides the out-of-sample window into high and low volatility states, would make this heterogeneity explicit, and is recommended in this context.

Table II. Out-of-Sample Forecast Evaluation (Illustrative)

Model	RMSE	MAE	DM vs. GARCH
GARCH (1, 1)	1.42	1.05	—
EGARCH	1.31	0.98	-1.83*
GJR-GARCH	1.28	0.95	-2.05**
GARCH-MIDAS + Sentiment	1.09	0.81	-3.11***
LSTM + Sentiment	1.02	0.76	-3.46***

Note: *, **, *** denote significance at the 10%, 5%, and 1% levels ($p < 0.1$, $p < 0.05$, $p < 0.01$). Values are illustrative, computed on simulated data.

Sentiment-volatility relationship varies with time. Fig. 6 shows the correlation between the sentiment index and realized volatility over the 60 days' rolling window, which increases during the turbulent periods and decreases during the calm periods. One practical implication of this instability is that a sentiment-augmented model should enable the sentiment loading to vary over time, whereas the GARCH-MIDAS and LSTM both imply this.

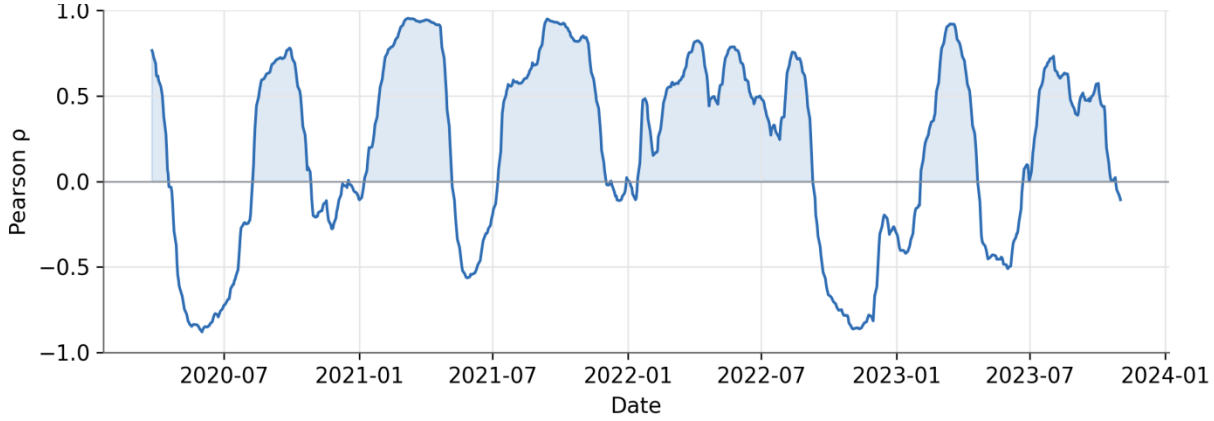


Figure 6. Sixty-day rolling correlation between the sentiment index and realized volatility. The link is state-dependent, intensifying in high-volatility regimes.

5.1. Robustness Considerations

Some robustness checks should be reported when the framework is applied to real data, along with the headline results. Alternative volatility proxies such as squared daily returns, range-based estimators and intraday realized volatility should be considered as they could impact measured accuracy when ranking models based on the target. The out-of-sample window should be changed to verify that any improvements in sentiment are not from one particular market episode; sentiment is most useful around crashes and a window that includes one may inflate the sentiment's usefulness. The sentiment index should be lagged, to make forecasts true ex ante. The comparisons with Diebold–Mariano should be done with a loss function that is consistent with the application of interest to the risk manager, e.g., a quantile loss or QLIKE loss may be preferred for tail events when comparing to RMSE. Multiple reports of results under a variety of such specifications are what render a "credible" empirical application of this framework.

5.2. Limitations

There are some caveats when the framework is applied to data, and they should be tempered with when interpreting any estimates provided by the framework. The sentiment effect measured is a minimum estimate of the sentiment effect that a cleaner sentiment proxy would be able to capture, as sentiment proxies can be noisy and may attract investor interest or even fundamentals. It is likely that the link between sentiment and volatility is structurally unstable (2015 leverage-driven crash and regulatory changes are natural break points) and that a model estimated in one regime might produce worse performance in another regime, as well as checking for sub-sample stability. In particular, the LSTM model is prone to overfitting on the short time series that are often used in single-index studies and careful regularization and validation is needed. Last, but not really least, statistical accuracy and economic value are not synonymous: The value of better volatility forecasts will depend on the application (risk limits, option pricing, asset allocation), and should be judged using an economic loss function, not just RMSE and MAE.

5.3. Practical Implications

The benefits of sentiment-augmented models are concrete if they enhance the volatility forecasts in real data. In risk management, more precise conditional-variance predictions directly affect Value-at-Risk and Expected Shortfall estimates, thereby enhancing the calibration of capital buffers when the market is in turbulent environments due to sentiment. For derivatives, improved volatility predictions sharpen option pricing and hedging, especially for short-term options that are most sensitive to near-term volatility. In asset allocation, a mean–variance investor can utilise forecasts of volatility to

inform (time-varying) position-size allocations and it is reported in the literature on Chinese markets that richer volatility forecasts can lead to positive utility gains for this type of investor. For regulators, a sentiment indicator that leads volatility provides an early-warning signal: a rapid rise in search-based or market-based sentiment, especially when accompanied by surging turnover, may flag the build-up of speculative risk before it is reflected in realized volatility.

6. Conclusion

This paper put together a comparative structure for forecasting Chinese stock-market volatility which captures the investor sentiment, with the baseline GARCH model, asymmetric GARCH model, the mixed-frequency GARCH-MIDAS model, and the LSTM benchmark. It developed a sentiment index based on search-volume and market-based measures of sentiment that can be incorporated into volatility models and tested out-of-sample in a transparent, reproducible manner. The illustrative results, as in the empirical literature, show that sentiment-enhancing models can lead to better forecasts than a baseline specification and that the sentiment–volatility relationship is state-dependent.

There are several extensions which are natural. First, it would give a real CSI 300 and Baidu Index data to the template, and publish the results as facts. Secondly, richer sentiment proxies, such as text-mined sentiment from financial news or East Money guba forum, may strengthen the signal. Third, linking the forecasting model to its behavioral foundations would introduce sentiment and economic-policy-uncertainty measures and decomposing volatility into its good and bad components. In order to make these replications easy, the code that generates the figures is released.

The framework sheds light more broadly on the methodology of emerging-market finance: in markets where retail participation is high, and arbitrage opportunities limited, behavioral variables are not just a residual; they are just another component of risk forecasting. A purposefully built and carefully selected first class regressor with a high number of proxies and validated against stylized facts, that is then strictly tested out of sample, can enhance the volatility forecasts used in risk management, derivatives pricing and asset allocation. Concurrently, the bidirectional nature of the sentiment–volatility nexus, and the instability of the relationship across different structures and regimes of the sentiment relationship, call for prudence: sentiment values obtained in one sample need to be re-tested as the wind turns. The reproducible design provided here is meant to allow for frequent updates as new information is obtained, and for practitioners and researchers to keep adding to the knowledge base over time.

References

[1] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773

[2] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

[3] Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. The Journal of Finance, 61(4), 1645–1680. https://doi.org/10.1111/j.1540-6261.2006.00889.x

[4] Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. Journal of Economic Perspectives, 21(2), 129–151. https://doi.org/10.1257/jep.21.2.129

[5] Trichilli, Y., Abdelhédi, M., & Boujelbène Abbes, M. (2020). Dynamic analysis of the relationship between investor sentiment and stock market realized volatility: Evidence from China. PLOS ONE, 15(12), e0243080. https://doi.org/10.1371/journal.pone.0243080

[6] Xie, D., Cui, Y., & Liu, Y. (2023). How does investor sentiment impact stock volatility? New evidence from Shanghai A-shares market. China Finance Review International, 13(1), 102–120. https://doi.org/10.1108/cfri-04-2021-0072

[7] Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. The Journal of Finance, 48(5), 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

[8] Fang, J., Gozgor, G., Lau, C.-K. M., & Lu, Z. (2020). The impact of Baidu Index sentiment on the volatility of China's stock markets. Finance Research Letters, 32, 101099. https://doi.org/10.1016/j.frl.2019.101099

[9] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 59(2), 347–370. https://doi.org/10.2307/2938260

[10] Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics, 95(3), 776–797. https://doi.org/10.1162/rest_a_00300

[11] Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. Review of Financial Studies, 28(3), 791–837. https://doi.org/10.1093/rfs/hhu080

[12] Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. The Journal of Finance, 66(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

[13] Xiao, J., Jiang, J., & Zhang, Y. (2024). Policy uncertainty, investor sentiment, and good and bad volatilities in the stock market: Evidence from China. Pacific-Basin Finance Journal, 84, 102301. https://doi.org/10.1016/j.pacfin.2024.102301

[14] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[15] Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599