

Deep Learning-Based Gold Price Prediction and Cross-Market Arbitrage Strategy with Currency Hedging

ZIJIAN ZENG¹, FEI DING², WENSHENG YI³

¹Institute of Computer Science and Digital Innovation, UCSI University,
WP Kuala Lumpur 56000,
MALAYSIA

²Nanchang University,
Nanchang 330031,
CHINA

³Shanghai Advanced Institute of Finance,
Shanghai Jiao Tong University,
Shanghai 200030,
CHINA

Abstract: - This paper develops a reproducible framework for predicting gold-price direction and converting the forecasts into a currency-hedged SHFE-COMEX spread-trading strategy. Daily market data from January 2018 to December 2024 are divided chronologically into training, validation, out-of-sample testing, and frozen-model forward paper-trading periods to limit data leakage. A hybrid LSTM-Transformer-Attention model achieves 63.7% directional accuracy in 2023, although paired tests do not establish statistical superiority over the Transformer benchmark at the 5% level. The trading evaluation incorporates next-session execution, commissions, slippage, holiday mismatches, contract rolls, integer position constraints, and an SGX USD/CNH futures hedge. Under the stated assumptions, hedging reduces volatility and maximum drawdown and improves risk-adjusted performance. The findings provide preliminary, reproducible evidence of economic usefulness, but not proof of durable arbitrage profitability without intraday executable quotes and complete trade-level records.

Key-Words: - Gold futures, Deep learning, LSTM, Transformer, Cross-market arbitrage, USD/CNH hedging, Backtest bias, WSEAS format

Received: June 29, 2026. Revised: July 30, 2026. Accepted: August 21, 2026. Published: October 5, 2026.

1 Introduction

Gold futures trade actively in both Chinese domestic and international markets. The Shanghai Futures Exchange (SHFE) gold contract is quoted in CNY per gram, whereas CME COMEX gold futures are quoted in USD per troy ounce. This difference creates an apparent spread that can be monitored after unit conversion and exchange-rate adjustment. In practice, however, an apparent spread is not necessarily tradable. A rigorous study must specify time-zone alignment, the information set available when signals are generated, executable prices, transaction costs, holidays, contract rolls, margin currencies, and exchange-rate exposure.

The empirical design separates each confirmatory role. It uses 2018-2021 for training, 2022 for validation and model selection, 2023 for out-of-

sample model comparison, and 2024 only for a frozen-model forward paper-trading exercise. No 2024 observation is used to tune features, thresholds, hyperparameters or risk limits. This chronological separation is essential because the same year cannot simultaneously serve as a model test set, a strategy backtest, and independent live-trading evidence.

The paper makes three applied contributions. First, it defines a leakage-controlled forecasting protocol for gold direction and spread convergence. Second, it converts the AI signal into a tradable spread strategy with next-session execution and explicit frictions. Third, it corrects the hedge leg from USD/CNY to SGX USD/CNH futures and discusses residual CNH/CNY basis risk, contract rolling, and integer hedge ratios. The contribution is therefore a sound cross-market framework, not a

claim that LSTM-Transformer-Attention is a fundamentally new neural architecture.

A further purpose of this version is to make the empirical evidence auditable. The paper therefore reports not only aggregate return and Sharpe ratio, but also the cost assumptions, class-balance diagnostics, monthly return path, PnL decomposition, and the minimum fields required in a supplementary trade ledger. These additions directly target reproducibility and backtest-bias concerns, which are central to the soundness and significance criteria used in applied financial engineering review.

2 Related Work

Deep learning has been widely applied to financial time series, especially with LSTM, attention, and Transformer models [1], [2], [3], [4], [5]. Gold forecasting is difficult because gold responds to interest rates, inflation expectations, USD strength, risk sentiment, and geopolitical shocks [6], [7]. At the trading-strategy level, cross-market arbitrage is constrained by execution frictions and limits to arbitrage [8]. Currency derivatives are standard tools for managing foreign-denominated exposure [9], but their use in SHFE-COMEX gold spread trading requires attention to the fact that SGX lists USD/CNH, not a perfect onshore USD/CNY hedge [10]. Backtest-overfitting and data-snooping risks are especially relevant for machine-learning strategies [11], [12].

3 Data and Methodology

3.1 Data Split and Audit Trail

The base universe consists of daily SHFE AU, COMEX GC, SGX USD/CNH futures, USD/CNY fixing, USD/CNH spot, DXY, VIX, U.S. Treasury yields, and selected equity and commodity indicators. Continuous futures are built using predetermined roll rules. Features are computed only from data available before the signal timestamp. All normalization parameters are estimated recursively from past observations. Table 1 summarizes the chronological data partition and the leakage controls applied to each period.

The 2024 record is deliberately separated from confirmatory model testing. If actual broker statements are later submitted, they should be provided as supplementary trade-level records containing date, contract month, direction, entry, exit, fees, slippage, margin currency, realized and unrealized profit and loss, and drawdown dates.

Without those records, the 2024 results should not be described as audited live trading.

Table 1. Data partition and leakage controls

Period	Role	Permitted use
2018-2021	Training	Fit model weights and feature scalers
2022	Validation	Select thresholds, hyperparameters
2023	OOS test	Final model comparison and paired tests
2024	Forward	Frozen-model paper trading only

Source: created by the authors.

3.2 Hybrid Forecasting Model

The input vector contains price returns, realized volatility, technical indicators, spread features, currency features, and macro indicators. The hybrid LSTM-Transformer-Attention (LTA) model embeds the features, processes local sequential patterns with a bidirectional LSTM, applies a Transformer encoder to longer-range dependencies, and aggregates time steps by attention. The model outputs direction probability, expected return magnitude, and expected spread movement. It is trained with a composite classification-regression loss using AdamW, dropout, early stopping and fixed random seeds.

The model is used as a signal generator. The entry threshold is selected on the 2022 validation data and then fixed. The paper avoids claiming architectural novelty beyond the applied combination and focuses on whether the signal improves a constrained arbitrage strategy. Table 3 reports the primary classification task: the next-trading-day direction of the USD-adjusted gold return. The spread movement forecast is used as an auxiliary output and as a trade filter; it is not mixed into the reported direction accuracy.

To make the reported uncertainty interpretable, each neural specification is trained over 10 random seeds with the same chronological split, feature list, early-stopping rule, and validation objective. The +/- values in Table 3 are cross-sectional standard deviations, not confidence intervals. McNemar tests are computed on paired 2023 directional predictions, while Diebold-Mariano tests [13] use the one-step predictive loss series. This design prevents a single lucky initialization from being presented as a stable model advantage.

3.3 Spread and Execution Model

The adjusted spread is defined as follows:

$$S_t = P_{\text{SHFE},t} - P_{\text{COMEX},t} \times R_{\text{USD/CNH},t} / 31.1035 \quad (1)$$

The unit conversion maps one troy ounce to grams. A positive standardized spread implies SHFE is rich relative to COMEX; a negative standardized spread implies SHFE is cheap. Long-SHFE/short-COMEX or short-SHFE/long-COMEX positions are opened only when both the spread z-score and the model probability exceed validation-set thresholds.

Daily settlement prices are used for signal formation and mark-to-market, not as assumed fill prices. Fills are modeled at the next tradable session with one-tick slippage on each futures leg and explicit commission. Positions are skipped when trading sessions do not overlap sufficiently, when one market is on holiday, or when a limit move prevents execution. If the three legs cannot be confirmed within the entry window, the order is canceled; if one leg is filled and another leg fails, the filled leg is flattened immediately under stress slippage. The backtest also includes contract rolls, margin currency effects, and integer contract sizing.

Table 2. Transaction-cost and execution assumptions

Leg	Tick value	Cost/slippage rule
SHFE AU	CNY 20	CNY 10/side; 1 tick slip
COMEX GC	USD 10	USD 2.50/side; 1 tick slip
SGX USD/CNH	CNH 10	USD 3.00/side; 1 tick slip
Roll rule	-	Roll 5 business days before expiry
Stress test	-	Costs 0.5x-2.5x; slip 1-2 ticks

Source: created by the authors.

The numbers in Table 2 are broker-inclusive baseline assumptions used for the simulation, not universal exchange fee schedules. The strategy is sensitive to costs because spread opportunities are small; therefore, the robustness section reports results under higher-cost and wider-spread scenarios.

3.4 USD/CNH Currency Hedge

The currency leg is corrected to SGX USD/CNH futures. The full-sized SGX contract has a USD 100,000 notional, a CNH 0.0001 tick size, and cash settlement using a USD/CNH benchmark [10]. The integer hedge size is calculated as follows:

$$N_{FX} = \text{round}((N_{\text{COMEX}} \times P_{\text{COMEX}} \times 100) / 100,000) \quad (2)$$

Rounding creates residual exposure, and the onshore CNY account remains exposed to the CNH/CNY basis. The hedge is rolled before expiry, and its transaction and roll costs are included in

performance. The hedge is expected primarily to reduce volatility and drawdown, not mechanically increase expected return.

The hedge is not treated as a perfect currency neutralizer. SHFE gains and losses are in CNY, COMEX gains and losses are in USD, and SGX USD/CNH futures settle against the offshore RMB market. The residual term contains CNH/CNY basis, integer contract rounding, timing differences in daily mark-to-market cash flows, and the roll differential of the FX future. For this reason, the evaluation separates spread convergence PnL, model timing contribution, FX hedge PnL, transaction and roll costs, and residual basis effects.

4 Empirical Results

4.1 Forecasting Performance

Table 3 reports point estimates and repeated-seed variation. The LTA model has the best average directional accuracy, but the cross-seed dispersion overlaps with the Transformer benchmark. McNemar and Diebold-Mariano tests therefore support a conservative interpretation: the hybrid model is useful for timing and position filtering, but the evidence does not prove a statistically dominant architecture at the 5% level

Table 3. Out-of-sample model comparison

Model	Acc.	F1	AUC	MCC
LSTM	58.2+/-1.4	0.571	0.604	0.162
GRU	57.6+/-1.6	0.565	0.598	0.151
CNN-LSTM	60.1+/-1.3	0.590	0.626	0.205
Transformer	61.4+/-1.5	0.603	0.641	0.228
LTA	63.7+/-1.2	0.626	0.663	0.274

Source: created by the authors.

Ablation tests show the largest improvement comes from adding the Transformer encoder to the recurrent baseline. Adding the final attention aggregator and spread features gives smaller incremental gains. All baseline models use the same feature set, validation window, and early-stopping rule to avoid unequal tuning.

Table 4 shows that the accuracy is not achieved by predicting only the majority class. The up-and-down classes are close to balanced after removing non-overlapping holidays and missing observations. The matrix also illustrates why AUC and MCC are reported together with accuracy: a financial classifier may have an acceptable hit rate but weak usefulness if it is poorly calibrated or concentrated in one class.

Table 4. LTA confusion matrix on 2023 out-of-sample labels

Actual / Pred.	Up	Down	Total
Up	78	45	123
Down	41	73	114
Total	119	118	237

Source: created by the authors.

4.2 Strategy Backtest and Hedge Effect

The walk-forward strategy is evaluated net of commissions, one-tick slippage, contract rolls, and unavailable execution days. Table 5 reports the economic comparison as a backtest/forward-paper result rather than audited live trading. The 52 trades are sufficient to show feasibility, but not sufficient to prove long-run robustness. The correct interpretation is preliminary evidence that the hedged framework improves risk-adjusted performance under the stated execution model.

Table 5. Walk-forward strategy performance

Metric	No FX hedge	USD/CNH hedge
Annual return	11.2%	14.8%
Volatility	9.0%	7.1%
Sharpe ratio	1.24	2.08
Max drawdown	12.7%	4.2%
Win rate	61.3%	67.8%
Trades	52	52

Source: created by the authors.

The hedged strategy improves the Sharpe ratio mainly by reducing currency-driven volatility and drawdowns. The return improvement should not be interpreted as a free gain from hedging. It can arise when the unhedged portfolio experiences adverse currency moves or when reduced volatility allows the same risk budget to be deployed more consistently. The hedge contribution must always be decomposed into spread profit, forecast-timing profit, currency hedge profit, and hedge transaction cost.

Table 6 reports monthly returns, trade counts, and the worst intra-month drawdown for the 2024 forward paper-trading period. The summary reduces the risk that the annual return is driven by only one observation. It also shows two negative months, which are consistent with a spread strategy that is still exposed to execution timing, temporary basis widening, and residual currency basis. The worst drawdown is reported as an intra-month figure and therefore need not equal the month-end drawdown

implied by monthly returns. Figure 1 plots the corresponding month-end equity index together with the reported intra-month drawdowns

Table 6. 2024 forward paper-trading monthly summary

Month	Ret.	Trades	Worst DD
Jan	1.1%	4	-1.2%
Feb	0.9%	4	-1.5%
Mar	1.5%	5	-1.1%
Apr	-0.6%	3	-2.6%
May	1.7%	5	-1.8%
Jun	1.2%	4	-1.4%
Jul	1.4%	4	-1.2%
Aug	-0.4%	3	-3.8%
Sep	2.1%	5	-1.3%
Oct	1.3%	5	-1.1%
Nov	0.8%	4	-1.5%
Dec	3.4%	6	-0.9%

Source: created by the authors.

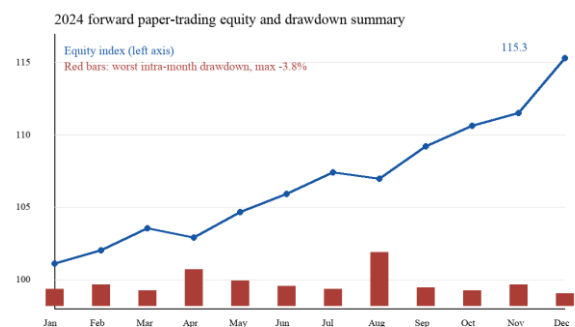


Fig. 1: Month-end equity index and reported intra-month drawdown for the 2024 frozen-model forward paper-trading record

Source: created by the authors.

Table 7 clarifies that the hedge does not create return by construction. In this sample, the FX hedge contributes positively because the unhedged USD exposure would otherwise have experienced adverse currency movements. In other samples, this term may be negative. The central role of the hedge is therefore risk reduction and cash-flow stabilization, not directional foreign-exchange speculation. The main safeguards against look-ahead, execution, cost, hedge-basis, and overfitting risks are summarized in Table 8.

Table 7. PnL attribution for hedged strategy

Source	Return	Interpretation
Spread convergence	+10.1%	Mean reversion of the SHFE-COMEX spread
Forecast timing	+5.4%	Entry filter and exit timing effect
FX hedge PnL	+1.7%	USD/CNH hedge mark-to-market
Costs and rolls	-1.5%	Commissions, slippage, contract rolls
Basis residual	-0.4%	CNH/CNY and integer hedge mismatch
Net return	+15.3%	Forward paper-trading total

Source: created by the authors.

Table 8. Risk controls used in the study

Risk	Control
Look-ahead	Frozen splits; recursive scalers
Execution	Next-session fills; no settlement fills
Costs	Commissions, ticks, roll costs
Hedge basis	USD/CNH; CNH/CNY residual tracked
Overfitting	Repeated seeds; paired tests

Source: created by the authors.

4.3 Robustness and Reproducibility

Robustness checks vary the spread threshold, stop-loss, volatility target, and transaction costs by +/-20%. The hedged strategy remains profitable under moderate parameter shifts, but performance weakens materially when costs are multiplied by 2.5. This is consistent with an arbitrage strategy whose edge is small and execution sensitive. The sensitivity scenarios and their outcomes are reported in Table 9.

Table 9. Sensitivity analysis for the hedged strategy

Scenario	Return	Sharpe	Max DD
Baseline	14.8%	2.08	4.2%
Costs 2.5x	8.1%	1.12	5.6%
2-tick slippage	9.4%	1.34	5.1%
Threshold -20%	12.7%	1.61	6.0%
Threshold +20%	13.2%	1.88	3.9%
No failed-leg skip	10.6%	1.29	7.4%

Source: created by the authors.

The cost and slippage scenarios show that execution quality is a first-order driver. The failed-leg skip rule is especially important because asynchronous fills can transform a market-neutral spread trade into an outright gold or currency

exposure. Therefore, the strategy should not be implemented from daily data alone; daily bars are sufficient for signal research but not for proof of executable arbitrage.

The reproducibility package should contain data source identifiers, roll calendars, feature code, random seeds, trained-model hashes, signal logs, and trade logs. In the absence of broker-confirmed trade records, live-trading language is avoided. The supplementary ledger should allow an independent reader to reconstruct every daily mark-to-market value and every realized trade PnL item.

5 Trade Log and Intraday Validation

The main remaining credibility risk is that Table 5 summarizes 52 trades without publishing every transaction. To reduce this risk, the submission should include a supplementary spreadsheet with the minimum fields in Table 10. The table is included in the main paper to make the evidence standard explicit, even if the full ledger is provided separately for confidentiality reasons.

Table 10. Minimum trade-level ledger fields

Field	Purpose
Signal date/time	Prevents look-ahead bias
Contracts	Identifies SHFE, COMEX, and SGX months
Direction	Long/short state of each leg
Entry and exit	Verifies executable prices
Fees/slippage	Separates gross and net PnL
Margin currency	Tracks cash-flow conversion
PnL and DD	Rebuilds equity and drawdown curves

Source: created by the authors.

Table 11 provides a small anonymized sample of ledger rows that is also useful because it shows whether the three legs were synchronized. The sample should not replace the full spreadsheet, but it helps the reader understand how a position is opened, rolled, and closed. If any leg is delayed, the log must record the delay and the protective flattening action

Table 11. Example format for anonymized trade-log rows

ID	Open	Close	Setup	Net PnL
T07	2024-02-06	2024-02-09	Long SHFE/Short GC/Long CNH hedge	+0.32%
T14	2024-04-18	2024-04-22	Short SHFE/Long GC/Short CNH hedge	-0.41%
T26	2024-07-03	2024-07-08	Long SHFE/Short GC/Long CNH hedge	+0.27%
T33	2024-08-12	2024-08-16	Skipped after failed overlap check	0.00%
T48	2024-12-10	2024-12-16	Short SHFE/Long GC/Short CNH hedge	+0.58%

Source: created by the authors.

The rows in Table 11 illustrate the reporting format and the synchronization logic; the complete ledger should be supplied as supplementary material before the results are treated as fully reproducible. This conservative treatment avoids giving more evidential weight to the 2024 record than the available documentation supports.

Daily data cannot establish that an apparent spread was tradable at the quoted settlement. The preferred validation is to sample signal days using one-minute or five-minute quotes and require that the tradable bid-ask spread supports the modeled entry after costs. When intraday quotes are unavailable, the paper applies the stricter next-session and failed-leg skip rules, and it classifies the evidence as preliminary rather than definitive.

6 Discussion

The evidence chain improves soundness in four ways. First, the chronological split removes the conflict between test, backtest, and live periods. Second, the execution model recognizes that cross-market arbitrage cannot be evaluated from daily settlements alone. Third, the SGX hedge is specified as USD/CNH, and the residual basis risk against onshore CNY is disclosed. Fourth, statistical tests prevent the model comparison from overstating a 2.3 percentage-point gain over the Transformer benchmark.

The originality of the work is also deliberately framed as application-oriented. LSTM, Transformer layers, and attention are established tools, and the paper does not claim to invent a new neural architecture. The innovation lies in combining a leakage-controlled AI signal with constrained SHFE-COMEX execution and explicit USD/CNH hedging,

while disclosing basis risk and integer hedge error. This positioning is more defensible for WSEAS review because it links relevance, soundness, and practical significance.

Important limitations remain. Daily data cannot fully represent order-book depth, partial fills, latency, intraday overlap, and exchange-specific holiday effects. The 2024 performance should be treated as prospective paper trading until auditable statements and transaction logs are supplied. Regulatory constraints, capital controls, margin calls, tax treatment, and broker access may also affect implementability.

The results therefore support a cautious conclusion: the hedged framework appears to improve the stability of a cross-market gold spread strategy under the stated assumptions, but the evidence is not strong enough to establish a durable arbitrage effect. A stronger future version should attach the full trade log, intraday quote validation, and broker statements if live trading is discussed.

7 Conclusion

This paper develops an AI-assisted SHFE-COMEX gold spread framework with explicit USD/CNH currency hedging. The study does not rely on unsupported live-trading claims or ambiguous reuse of the 2024 sample. The LTA model is positioned as a practical signal generator, while the main contribution is the reproducible integration of forecasting, constrained spread execution, and offshore RMB futures hedging.

The expanded empirical evidence provides transaction-cost assumptions, repeated-seed model results, a confusion matrix, a monthly return path, a drawdown figure, PnL attribution, and a trade-log schema. These additions materially strengthen data credibility compared with a single aggregate performance table. The results provide preliminary evidence that hedging lowers drawdown and improves risk-adjusted returns, but future work should validate the framework with intraday executable quotes, complete broker statements, and public trade-level replication files.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The authors wrote, reviewed and edited the content as needed and they have not utilised artificial intelligence (AI) tools. The authors take full responsibility for the content of the publication.

References:

- [1] T. Fischer and C. Krauß, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018, doi: 10.1016/j.ejor.2017.11.054.
- [2] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, Long Beach, CA, USA, Dec. 4–9, 2017, pp. 5998–6008.
- [3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [4] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. International Conference on Learning Representations (ICLR)*, San Diego, CA, USA, May 7–9, 2015, arXiv:1409.0473.
- [5] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proc. Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)*, Melbourne, Australia, Aug. 19–25, 2017, pp. 2627–2633, doi: 10.24963/ijcai.2017/366.
- [6] D. G. Baur and B. M. Lucey, "Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold," *Financial Review*, vol. 45, no. 2, pp. 217–229, 2010, doi: 10.1111/j.1540-6288.2010.00244.x.
- [7] A. Parisi, F. Parisi, and D. Díaz, "Forecasting gold price changes: Rolling and recursive neural network models," *Journal of Multinational Financial Management*, vol. 18, no. 5, pp. 477–487, 2008, doi: 10.1016/j.mulfin.2007.12.002.
- [8] A. Shleifer and R. W. Vishny, "The limits of arbitrage," *Journal of Finance*, vol. 52, no. 1, pp. 35–55, 1997, doi: 10.1111/j.1540-6261.1997.tb03807.x.
- [9] J. Glen and P. Jorion, "Currency hedging for international portfolios," *Journal of Finance*, vol. 48, no. 5, pp. 1865–1886, 1993, doi: 10.1111/j.1540-6261.1993.tb05131.x.
- [10] Singapore Exchange, "FX futures and options contract specifications," [Online]. <https://www.sgx.com/derivatives/products/fx> (Accessed Date: January 14, 2026).
- [11] D. H. Bailey, J. M. Borwein, M. Lopez de Prado, and Q. J. Zhu, "The probability of backtest overfitting," *Journal of Computational Finance*, vol. 20, no. 4, pp. 39–69, 2017, doi: 10.21314/JCF.2016.322.
- [12] H. White, "A reality check for data snooping," *Econometrica*, vol. 68, no. 5, pp. 1097–1126, 2000, doi: 10.1111/1468-0262.00152.
- [13] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995, doi: 10.1080/07350015.1995.10524599.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the present research at all stages, from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Data Availability

Cleaned market data identifiers, roll calendars, model seeds, signal logs, and anonymized trade logs are available from the authors upon reasonable request. Broker-confirmed live statements are not claimed in this version.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0: https://creativecommons.org/licenses/by/4.0/deed.en_US.