Dual-Stream Deep Reinforcement Learning for Pairs Trading in Chinese Commodity Futures Markets

Guanrou Deng
Institute of Finance Technology,
University College London
Goldman Sachs
London, United Kingdom
guanrou.deng.21@ucl.ac.uk

Xinyu Lin
Institute of Finance Technology,
University College London
London, United Kingdom
xinyu.lin.23@ucl.ac.uk

Li Zhang*
Institute of Finance Technology,
University College London
London, United Kingdom
ucesl07@ucl.ac.uk

Abstract

Pairs trading, a market-neutral statistical arbitrage strategy, has gained renewed attention with advances in Deep Reinforcement Learning (RL). However, most RL-based approaches rely solely on endogenous price signals, overlooking exogenous market drivers. This limitation is particularly critical in Chinese commodity futures markets, where trading is shaped by domestic macroeconomic conditions and international commodity prices. We propose a dual-stream RL framework for pairs trading that integrates endogenous market information with exogenous features, including Chinese macroeconomic indicators and U.S. commodity futures. Built upon a Deep Q-Network (DQN) with an External Stream Network (ESN), our architecture models heterogeneous data sources within a unified decision-making process. Using ten years of real-world Chinese commodity futures data (2014–2023), we show that the framework significantly outperforms traditional statistical methods and standard RL baselines in profitability and risk-adjusted returns. Ablation studies further highlight the value of external signals and validate the dual-stream design's effectiveness in dynamic trading environments.

CCS Concepts

• **Computing methodologies** → **Sequential decision making**;
• **Applied computing** → **Economics**.

Keywords

Pair trading, Chinese commodity futures, Deep Q-Network, Dual-stream framework

ACM Reference Format:
Guanrou Deng, Xinyu Lin, and Li Zhang. 2025. Dual-Stream Deep Reinforcement Learning for Pairs Trading in Chinese Commodity Futures Markets. In *Proceedings of (ICAIF 25).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding author

ICAIF 25, Singapore

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The rapid evolution of financial markets has led to substantial progress in trading strategies, especially in the area of statistical arbitrage. Pairs trading stands out as a significant strategy due to its market-neutral characteristics and its capacity to exploit relative price movements between co-moving assets Elliott et al. [11]. In recent years, this domain has experienced a methodological shift, moving beyond traditional mean-reversion techniques toward data-driven strategies powered by deep learning and reinforcement learning. Approaches incorporating Long Short-Term Memory (LSTM) networks, Transformers, Convolutional Neural Networks (CNNs), Temporal Convolutional Networks (TCNs), and reinforcement learning have demonstrated considerable potential in capturing complex, nonlinear dependencies in financial time series [3, 6, 16, 17, 19, 33, 40, 44]. These advances enable dynamic and predictive modeling of market behavior, offering more adaptive and robust trading strategies across both domestic and international markets.

Over the past two decades, the Chinese commodity futures market has rapidly expanded into a major global player, driven by China's position as a leading producer and consumer of metals, energy, and agricultural products [31, 41]. With broader trading instruments, growing foreign participation, and a distinct regulatory environment, its pricing dynamics often diverge from Western markets [20, 36, 39]. As the market matures and integrates globally, price behavior is increasingly influenced by both domestic macroeconomic factors (e.g., inflation, industrial output, interest rates) and international forces (e.g., exchange rates, global commodity prices) [8, 43, 45]. Designing effective pairs trading strategies thus requires incorporating macroeconomic and cross-market signals. However, existing approaches typically rely on price-only statistical models and are mainly developed for equities or more mature commodity markets [12, 22, 33], leaving few studies tailored to China's futures market. Integrating macroeconomic and cross-market information into pairs trading remains particularly challenging due to data heterogeneity—differences in scale, frequency, and semantics between inputs—and environmental uncertainty caused by non-stationary market dynamics and unpredictable spread deviations.

To address these challenges, we propose a novel dual-stream framework based on the Deep Q-Network (DQN) algorithms [27]. The first stream: Internal Stream Network(ISN), models endogenous market dynamics by capturing temporal dependencies in historical spread data using Gated Recurrent Units (GRUs), alongside technical indicators processed by a Multi-Layer Perceptron (MLP).

The second stream: External Stream Network(ESN), encodes exogenous signals, such as domestic macroeconomic variables and international commodity prices. This dual-stream design enables the agent to form a unified state representation that embeds both types of information, facilitating adaptive decision-making under complex and changing market conditions. We also develop a robust action space tailored for real-world trading, comprising four discrete actions: *hold*, *long*, *short*, and *exit*. The *exit* action encompasses multiple exit triggers, including *stop-loss*, *capital depletion*, and *time-based exits*, to better reflect practical trading constraints and risk management needs.

We evaluate the proposed framework on real-world data from the Chinese commodity futures market (2014–2023), augmented with Chinese macroeconomic indicators and U.S. commodity futures prices. The effectiveness of our strategy is assessed using standard financial performance metrics such as annualized return, Sharpe ratio, and maximum drawdown. Results show that our framework consistently outperforms baseline methods in both returns and risk-adjusted performance. We further analyse the learned trading signals for interpretability and conduct ablation studies to assess the contributions of macroeconomic and international features, as well as the role of the ESN module in enhancing performance.

In summary, our contributions are as follows:

- We propose a novel dual-stream reinforcement learning framework for pair trading that integrates both endogenous market signals and exogenous contextual features, including Chinese macroeconomic indicators and U.S. commodity futures prices.
- We design a deep Q-network (DQN)-based trading agent with a robust and risk-aware action space, incorporating realistic exit mechanisms such as *stop-loss*, *capital depletion*, and *time-based exits* to better reflect practical trading constraints.
- Extensive real-world experiments validate the effectiveness of our framework and highlight the role of macroeconomic and international features in responding to market dynamics.

2 Related works

2.1 Pairs Trading Methodology

The literature on pairs trading has evolved to include two main methodologies: the distance approach and the cointegration approach [23, 38]. The distance approach selects pairs based on the historical closeness of normalized price series. Gatev et al. [15] shows that matching stocks by minimum distance can yield annualized excess returns of up to 11%, while Do and Faff [10] uses a nonparametric squared-difference metric to identify mean-reverting pairs. The cointegration approach uses statistical tests to detect pairs with a stable long-run equilibrium despite short-term deviations. Lin et al. [24] optimizes entry and exit thresholds for cointegration-based strategies, improving profitability, while comparative studies examine distance versus cointegration approaches in U.S. equity markets [29, 35].

Recently, machine learning has gained attention in pairs trading. Sarmento and Horta [33] combines OPTICS clustering with forecasting models (LSTM, ARMA), and Chang et al. [5] applies cointegration with LSTM to improve prediction accuracy for aggressive stocks. Reinforcement learning (RL) is particularly suited to the sequential nature of trading. Kim and Kim [22] proposes a DQN-based strategy for dynamic trading and stop-loss thresholds, outperforming fixed rules. Brim [4] extends this with a Double DQN to better capture mean reversion, while Kim et al. [21] introduces HDRL-Trader, a hybrid model with two RL networks achieving strong results on S&P 500 stock pairs.

Building on these works, we explore RL approaches and extend them to Chinese commodity futures pairs trading.

2.2 Pairs trading for Chinese Commodity Futures market

Researchers have explored the Chinese commodity futures market from various perspectives. Yang et al. [42], for instance, compares the profitability of different pair selection and spread trading methods using a comprehensive dataset covering multiple Chinese exchanges. Fernandez-Perez et al. [13] applies a time-series pairs trading model on Chinese and international commodity futures, the study finds robust excess returns, especially in metal and gold futures, outperforming traditional strategies.

However, most existing models focus solely on commodity price dynamics, overlooking the critical influence of macroeconomic indicators and international markets. Borensztein and Reinhart [2] enhances traditional demand-based models by incorporating supply-side and global demand factors, resulting in improved forecasting performance. Hess et al. [18] further demonstrates that commodity futures respond strongly to U.S. macroeconomic announcements, particularly during recessions, underscoring their role as hedges against economic shocks.

Beyond domestic macroeconomic effects, Chinese commodity futures are also shaped by developments in U.S. markets. Fung et al. [14] and Liu and An [25] find significant information spillovers from U.S. to Chinese markets, especially for commodities like copper and soybeans, with the U.S. market often playing a leading role. Building on this evidence, our study incorporates key U.S. commodity futures and Chinese macroeconomic indicators to enhance the predictive accuracy of RL-based pair trading strategies in China's futures market.

3 Proposed Methodology

Building on prior research, we propose a reinforcement learning (RL) framework for pairs trading in the Chinese commodity futures market. Tradable pairs are first identified using cointegration analysis. A deep RL agent is then trained to make trading decisions based on state representations combining price spread features with macroeconomic and international commodity indicators.

3.1 Problem Formulation

We model the pairs trading task in the Chinese commodity futures market as a discrete-time *Markov Decision Process (MDP)* [1], represented by $(\mathcal{S}, \mathcal{A}, \mathcal{T}, r, \gamma)$. The state space $\mathcal{S}$ includes endogenous market variables (e.g., spread, technical indicators) and exogenous features (macroeconomic indicators and U.S. market data). The action space $\mathcal{A}$ consists of discrete trading decisions: *long*, *short*, *hold*, or *exit*. The transition function $\mathcal{T}$ captures stochastic market evolution, and the reward r represents trading profit or loss, accounting for transaction costs. The discount factor $\gamma \in [0, 1)$ balances immediate and future rewards.

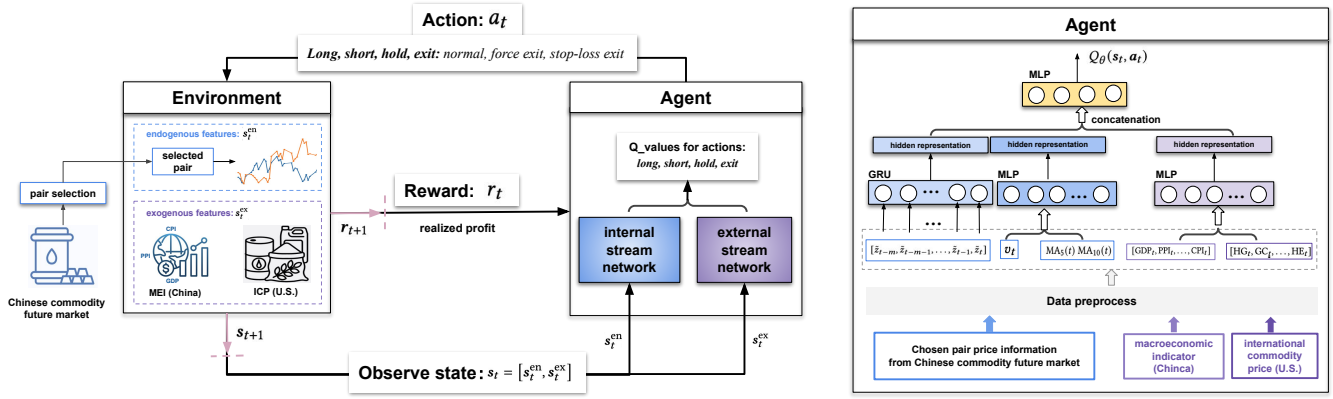


Figure 1: The overall framework of our dual-stream Deep RL framework for pairs trading in Chinese commodity futures markets.

At each timestep t, the agent observes $\boldsymbol{s}_t \in \mathcal{S}$ and selects $a_t \in \mathcal{A}$ according to a stochastic policy $\pi : \mathcal{S} \to \Delta(\mathcal{A})$. The environment transitions to $\boldsymbol{s}_{t+1} \sim \mathcal{T}(\boldsymbol{s}_t, \boldsymbol{a}_t)$, yielding reward $r_t = r(\boldsymbol{s}_t, \boldsymbol{a}_t)$. The objective is to learn an optimal policy π^* that maximizes the expected cumulative discounted reward:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_\pi \left[\sum_{t=0}^{T} \gamma^t r(\boldsymbol{s}_t, \boldsymbol{a}_t) \right].$$

This formulation unifies heterogeneous market signals and provides a principled basis for learning adaptive trading strategies via reinforcement learning.

3.2 Pair selection

The selection of commodity futures pairs is conducted using the cointegration methodology. Cointegration emphasizes the stationarity of the spread between assets, aiming to identify pairs of commodity futures that exhibit a stable long-term equilibrium relationship [28, 30]. We use the price spread between each pair of commodity futures as input to test for cointegration

$$P_{i,t} = \beta P_{j,t} + \varepsilon_t, \tag{1}$$

where $P_{i,t}$ and $P_{j,t}$ are the two commodity futures prices being tested for cointegration. Then the residuals (ε_t) are tested for stationarity using the Augmented Dickey-Fuller (ADF) test [7].

Based on the resulting p-values from the ADF test, we retain the pairs with a p-value <0.01, indicating evidence against the null hypothesis that the series are not cointegrated. Therefore, the pairs that pass the ADF test are selected for further trading strategies.

3.3 Environment for pair trading

3.3.1 State Space. As introduced above, we define the state space $\mathcal{S}$ as the set of all information available to the agent at each decision step. Each state $\boldsymbol{s}_t \in \mathcal{S}$ at time t comprises two components: endogenous pricing features $\boldsymbol{s}_t^{\text{en}}$ and exogenous contextual features $\boldsymbol{s}_t^{\text{ex}}$. Formally, the full state can be written as:

$$\boldsymbol{s}_t = \left[\boldsymbol{s}_t^{\text{en}}, \boldsymbol{s}_t^{\text{ex}} \right]. \tag{2}$$

On the one hand, endogenous pricing features capture mean-reversion behavior and short-term dynamics of the trading pair. These include a rolling window of standardized spreads, spread returns, moving averages (MA), and the agent's current position [32, 37]. Standardization of spreads is widely used in statistical arbitrage to measure deviations from long-term equilibrium and identify trading signals. Moving averages (MA5, MA10) are computed to smooth noise and capture short- to medium-term trends:

$$\mathrm{MA}_k(t) = \frac{1}{k} \sum_{i=0}^{k-1} \tilde{z}_{t-i}, \quad k \in \{5, 10\}. \tag{3}$$

The full endogenous feature vector is:

$$\boldsymbol{s}_t^{\text{en}} = \left[\tilde{z}_{t-m}, \ldots, \tilde{z}_t,\ v_t,\ \mathrm{MA}_5(t),\ \mathrm{MA}_{10}(t),\ p_t \right], \tag{4}$$

where v_t is the spread return and p_t is the current position related to action a_t (see Section 3.3.2).

On the other hand, the exogenous contextual features in our state space include a set of Chinese macroeconomic indicators and international commodity prices from the U.S. market. Specifically, for the macroeconomic indicators, we incorporate China's GDP growth rate, Producer Price Index (PPI), Consumer Price Index (CPI), Core CPI (CCPI), and the 5-year and 1-year Loan Prime Rates (LPR-5, LPR-1). These indicators jointly reflect the domestic economic environment and monetary policy stance: GDP captures aggregate demand and economic growth; PPI and CPI reflect inflationary pressures from the production and consumption sides; CCPI offers a more stable measure of inflation by excluding volatile items and LPRs are key policy rates that influence credit costs and investment behavior. The macroeconomic vector is expressed as:

$$\boldsymbol{m}_t = \left[\mathrm{GDP}_t,\ \mathrm{PPI}_t,\ \mathrm{CPI}_t,\ \mathrm{CCPI}_t,\ \mathrm{LPR}_t^{(1)},\ \mathrm{LPR}_t^{(5)} \right] \in \mathbb{R}^6. \tag{5}$$

In addition, we incorporate prices of key U.S.-traded commodity futures as exogenous signals. These include benchmark contracts for metals (copper HG, gold GC, silver SI), energy (crude oil CL, natural gas NG), and agricultural products (soybeans ZS, soybean oil ZL, corn ZC, and lean hogs HE). Therefore, we define the international

commodity future price vector as:

$$\boldsymbol{u}_t = [\mathrm{HG}_t,\ \mathrm{GC}_t,\ \mathrm{SI}_t,\ \mathrm{CL}_t,\ \mathrm{NG}_t,\ \mathrm{ZS}_t,\ \mathrm{ZL}_t,\ \mathrm{ZC}_t,\ \mathrm{HE}_t] \in \mathbb{R}^9. \tag{6}$$

Then, the full exogenous feature vector is then written as:

$$\boldsymbol{s}_t^{\mathrm{ex}} = [\boldsymbol{m}_t,\ \boldsymbol{u}_t] \in \mathbb{R}^{15}, \tag{7}$$

combining both macroeconomic conditions and global commodity price signals. This design enables the model to respond to international price movements and policy-driven economic changes, improving its capacity for identifying robust trading opportunities.

3.3.2 Action Space. To make the agent's decision process both more efficient and more aligned with real-world trading conditions, we design the trading action as a categorical probability vector $\boldsymbol{a}_t \in \mathbb{R}^4$. The agent executes the action with the maximum probability. Each action determines the trading position p_t for the selected pair, with unit position held or closed. At time t, actions can be written as

$$a_{it} \in \{hold, long, short, exit\}, i \in [0, 1, 2, 3], \tag{8}$$

where *hold* means take no position, *long* means long an an undervalued stock i and short an overvalued stock j, *short* denotes the inverse of the long action—shorting the undervalued asset while taking a long position in the overvalued one.

In particular, the *exit* action is context-dependent and may occur through four mechanisms. A *normal exit* happens when the agent closes a position based on its learned policy. A *timeout exit* is triggered when the maximum holding duration is reached. A *stop-loss exit* occurs if the unrealized return drops below a threshold defined by the spread's historical standard deviation. Finally, a *capital constraint exit* is used to terminate the episode when the portfolio value falls to zero or below.

3.3.3 Reward. The reward at each time step is defined as the realized profit and loss (PnL) from executing a trading action, net of transaction costs. Specifically, the PnL is computed based on the change in spread value between entry and exit, adjusted by the direction of the position (long or short). Let $e_{t_{\mathrm{entry}}}$ and $e_{t_{\mathrm{exit}}}$ denote the spread at the entry and exit time, respectively, and let $d \in \{1, -1\}$ represent the direction of the position (1 for long, -1 for short). The gross PnL is given by:

$$\mathrm{PnL}_t = d \cdot \left(e_{t_{\mathrm{exit}}} - e_{t_{\mathrm{entry}}}\right). \tag{9}$$

A round-trip transaction incurs transaction costs on both entry and exit. Let κ denote the transaction cost rate. The total transaction cost is:

$$\mathrm{Cost}_t = \kappa \cdot \left(e_{t_{\mathrm{exit}}} + e_{t_{\mathrm{entry}}}\right). \tag{10}$$

The final reward is thus defined as:

$$r_t = \mathrm{PnL}_t - \mathrm{Cost}_t. \tag{11}$$

This reward structure encourages the agent to identify profitable mean-reverting opportunities while penalizing over-trading and noise-driven actions, thereby promoting stable and cost-aware trading behavior.

3.3.4 DQN Agent Architecture and Learning Framework. Our reinforcement learning framework builds on the Deep Q-Network (DQN) algorithm [27], where the state includes both endogenous pricing features $\boldsymbol{s}_t^{\mathrm{en}}$ and exogenous contextual features $\boldsymbol{s}_t^{\mathrm{ex}}$, capturing Chinese macroeconomic indicators and U.S. commodity prices. To process these heterogeneous inputs, we design a dual-stream neural architecture with Q-value modulation, enabling the agent to jointly reason over internal market signals and external information.

The first stream: the internal stream network, encodes endogenous market features, denoted by $\boldsymbol{s}_t^{\mathrm{en}} \in \mathbb{R}^d$, which capture the internal dynamics of commodity future prices. Specifically, the feature vector comprises a sequence of historical spread values , volume indicators, and technical signals. To model temporal dependencies within this sequence, we employ a Gated Recurrent Unit (GRU) network [9] to get the hidden representation:

$$\boldsymbol{h}_t^{\mathrm{en}} = \mathrm{GRU}_{\mathrm{en}}(\boldsymbol{s}_{1:t}^{\mathrm{en}}). \tag{12}$$

In parallel, the second stream: the external stream network, processes exogenous signals $\boldsymbol{s}_t^{\mathrm{ex}}$, which include macroeconomic indicators and international commodity prices. These features provide contextual information on external market conditions and potential regime shifts. A temporal encoder f_{ex}, implemented as a feedforward or recurrent model, transforms the input into a latent representation:

$$\boldsymbol{h}_t^{\mathrm{ex}} = f_{\mathrm{ex}}(\boldsymbol{s}_{1:t}^{\mathrm{ex}}). \tag{13}$$

The resulting hidden representations from both streams are concatenated and passed through a fully connected network f to estimate Q-values for all available actions:

$$Q_\theta(\boldsymbol{s}_t, \boldsymbol{a}_t) = f\left(\left[\boldsymbol{h}_t^{\mathrm{en}}, \boldsymbol{h}_t^{\mathrm{ex}}\right]\right), \tag{14}$$

where $\theta = \{\theta_{\mathrm{GRU}}, \theta_{f_{\mathrm{ex}}}, \theta_f\}$ denotes the full set of learnable parameters. The final output layer produces a Q-value for each discrete action $a \in \mathcal{A}$, where $|\mathcal{A}|$ is the cardinality of the action space:

$$Q_\theta(\boldsymbol{s}_t, a), \quad \forall a \in \mathcal{A}.$$

3.4 Training Procedure

The agent Q_θ aims to learn an action-value function that accurately estimates the expected cumulative return for each trading decision. To achieve this, it is trained by minimising the temporal difference loss, which measures the discrepancy between the current Q-value prediction and a one-step bootstrapped target:

$$\mathcal{L}(\theta) = \mathbb{E}_{(\boldsymbol{s}_t, \boldsymbol{a}_t, r_t, \boldsymbol{s}_{t+1}) \sim \mathcal{D}}\left[\left(Q_\theta(\boldsymbol{s}_t, \boldsymbol{a}_t) - Q_\theta^*(\boldsymbol{s}_{t+1}, \boldsymbol{a})\right)^2\right], \tag{15}$$

where

$$Q_\theta^* = r_t + \gamma \max_{a' \in \mathcal{A}} Q(\boldsymbol{s}_{t+1}, a'; \theta^-),$$

$\gamma \in [0, 1]$ is the discount factor and θ^- are parameters of a periodically updated target network. The discrete action set $\mathcal{A}$ includes *hold*, *long*, *short*, and *exit*.

To improve efficiency and stability, we use experience replay: past transitions $(\boldsymbol{s}_t, \boldsymbol{a}_t, r_t, \boldsymbol{s}_{t+1})$ are stored in a buffer $\mathcal{D}$ and sampled uniformly for mini-batch updates. Parameters θ are optimized via stochastic gradient descent and backpropagation. In the meantime, we use the ϵ-greedy policy that balances exploration and

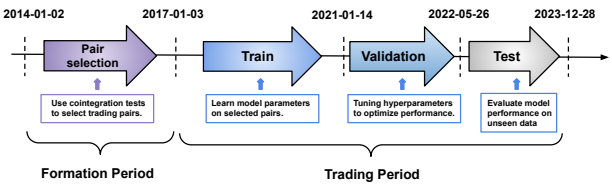


Figure 2: Data Timeline

exploitation, selecting a random action with probability ϵ decays exponentially each episode:

$$\epsilon \leftarrow \max(\epsilon_{\min}, \epsilon \cdot \epsilon_{\text{decay}}),$$

and choosing greedily during inference:

$$a_t = \arg\max_a Q(s_t, a).$$

4 Experiments

4.1 Experimental Setup

4.1.1 Data. We construct our dataset from the Wind and Yahoo Finance platforms, covering both commodity futures prices and macroeconomic indicators relevant to the Chinese market.

First, Chinese commodity futures data and macroeconomic indicators are sourced from the Wind database. The futures data include actively traded contracts on the Dalian Commodity Exchange (DCE) and Shanghai Futures Exchange (SHFE), with daily settlement prices and trading volumes across metals, energy, and agriculture. For the macroeconomic indicators used in this study, GDP is updated quarterly, while PPI, CPI, CCPI, and LPR are released monthly. To match the daily futures data, we forward-fill each indicator so that its latest available value is used until the next release. This ensures the model has access to the same information a trader would observe each day. To capture international influences, U.S. commodity futures prices (Section 3.3.1) are obtained from Yahoo Finance.

The dataset spans the period from January 2014 to December 2023. The first three years (2014–2016) are used as a formation period to identify stable and economically meaningful commodity pairs based on historical price relationships. The remaining seven years (2017–2023) are split into training, validation, and testing sets for model development and performance evaluation as illustrated in Figure 2.

4.1.2 Baselines and experimental setting. We compare our model with several baseline strategies commonly used in the literature for pair trading, ranging from simple heuristics to advanced learning-based approaches. These baselines include:

- Buy and Hold (B&H): Buys both assets on the first day of the test period and sells them on the last day.
- Statistical arbitrage trading strategy [15] (SA): Opens a position when the normalized spread deviates beyond a threshold (e.g., ±2) and closes it upon mean reversion.
- DQN [22]: Learns discrete trading and stop-loss boundaries from spread features using a ReLU network with a softmax output.
- Double DQN [3]: Extends DQN with double estimators to reduce Q-value overestimation.
- Our Method: Builds on DQN to learn policies from spread features and exogenous signals via a dual-stream architecture, capturing both price dynamics and contextual market information.

For all RL-based methods, hyperparameters are tuned via grid search (hidden units $\{8, 16, 32, 64\}$, learning rates $\{0.001, 0.005, 0.01\}$), selecting the configuration with the highest validation cumulative return. The initial capital is 100,000 with a 0.1% transaction cost, a maximum holding period of 30 days, and a stop-loss of 2.0 standard deviations. We use ϵ-greedy exploration, decaying ϵ from 1.0 to 0.01 with factor 0.995, a discount factor $\gamma = 0.9$, and experience replay with buffer size 2000 and mini-batch size 32.

4.1.3 Evaluation. We evaluate each trading strategy using five standard financial metrics: compound annual return (CAR), annualized return (AR), Sharpe ratio [34], maximum drawdown (MDD) [26], and total holding period. These metrics, widely used in financial performance evaluation [15, 23], jointly capture profitability, risk-adjusted performance, downside risk, and market exposure. CAR and AR measure capital growth, the Sharpe ratio compares excess return to volatility, MDD quantifies the largest observed loss from a peak, and the holding period reflects trading activity time length.

4.2 Experimental Results

4.2.1 Pair Selection Results. The final set of tradable pairs is selected using the cointegration-based methodology outlined in Section 3.2. Although the selection is grounded in statistical testing, the identified pairs also exhibit strong economic rationale, reflecting coherent linkages across commodity sectors.

Table 1 summarizes the five selected pairs along with their associated commodity names and cointegration p-values, all of which fall below the 1% significance level. Specifically, CU.SHF and AL.SHF represent industrial metals shaped by infrastructure investment and cyclical demand from the manufacturing sector. B.DCE and M.DCE are linked through the energy and chemical industries, with crude oil often serving as a common cost driver. RU.SHF appears in multiple pairs, such as with P.DCE (PTA, p = 0.0077) and PP.DCE (polypropylene, p = 0.0066), reflecting its integration with the petrochemical supply chain and synthetic materials manufacturing. The pair RU.SHF and FU.SHF (fuel oil, p = 0.0065) further highlights a connection through transportation demand and oil market dynamics.

Overall, these pairs not only pass statistical cointegration thresholds but also align with intuitive economic relationships in the Chinese commodity futures market, making them suitable candidates for strategy development.

Table 1: Selected Cointegrated Commodity Futures Pairs

Ticker Pair	Commodity Names	p-value
CU.SHF & AL.SHF	Copper & Aluminum	0.0012
B.DCE & M.DCE	Bitumen & Methanol	0.0034
RU.SHF & FU.SHF	Natural Rubber & Fuel Oil	0.0065
RU.SHF & PP.DCE	Natural Rubber & Polypropylene	0.0066
RU.SHF & P.DCE	Natural Rubber & PTA	0.0077

4.2.2 Main results. Table 2 presents a detailed comparison across traditional benchmark models and different RL approaches across five selected commodity futures pairs. In all cases, RL models significantly outperform traditional strategies such as Buy-and-Hold (B&H) and SA, both in terms of return and risk-adjusted metrics. For instance, in the CU.SHF & AL.SHF pair, our proposed DRL model achieves an annualized return of 294.6% and a Sharpe ratio of 4.50, highlighting its ability to effectively capture the strong co-movement in industrial metals. It is worth noting that the exceptionally high return for this pair is largely driven by its inherent profitability, as evidenced by its benchmark performance—under both SA and Buy-and-Hold strategies, the returns are at least two to three times higher than those of other pairs. Overall, this emphasises both the model's ability to capture highly rewarding opportunities.

Additionally, in the B.DCE & M.DCE pair, our model outperforms others with a Sharpe ratio exceeding 7.1, reflecting solid risk-adjusted performance even under modest price dynamics. For pairs related to rubbers, our method maintains robust profitability and stability. The RU.SHF & FU.SHF pair shows annualized returns above 168.3% with a Sharpe ratio of 6.93, while RU.SHF & PP.DCE reaches a Sharpe ratio of 7.48. These results suggest that our model is well-suited to handling commodities with shared petrochemical dependencies.

Finally, to demonstrate the model's overall effectiveness, we compute the average performance across all five pairs, as shown in Table 3. While some pairs individually achieve high returns, our proposed method still delivers the best average performance overall, with a compound annual return of 4.38 and a Sharpe ratio of 6.18. This reinforces the model's generalizability and effectiveness across a diverse set of trading environments.

4.2.3 Trading signal analysis. We use the CU.SHF & AL.SHF pair to illustrate the trading signal results generated by our strategy. As shown in Figure 3, the blue line plots the spread between the two futures contracts, while the overlaid markers represent the model's trading decisions—green triangles for long entries, red inverted triangles for short entries, and orange circles for exit points.

The figure demonstrates that our agent is able to effectively identify key turning points in the spread's movement. For example, during the period between April and May 2023, the spread exhibits a downward trend, and the agent correctly responds by initiating short positions. Conversely, from June to July 2023, the spread trends upward, leading the agent to enter long positions. This dynamic adjustment showcases the agent's ability to capture mean-reverting behavior and adapt its strategy based on changing market conditions.

4.2.4 Effective of macro and us price. This subsection investigates the performance impact of incorporating two categories of features into our pair trading framework: (1) domestic macroeconomic indicators, and (2) U.S. commodity futures. We analyze their effects on three representative futures pairs: CU.SHF & AL.SHF (Industrial Metals), B.DCE & M.DCE (Energy and Chemicals), and RU.SHF & P.DCE (Consumer-Linked Industrials). For each pair, we augment the model by introducing one additional feature at a time, and measure the relative improvement in annual return over the baseline.

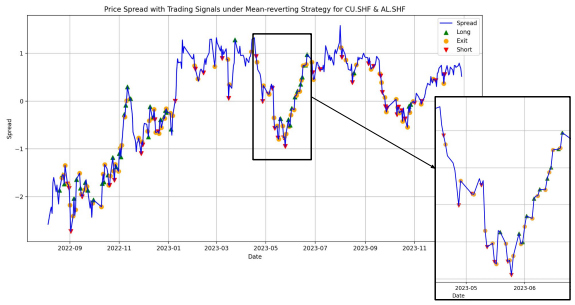
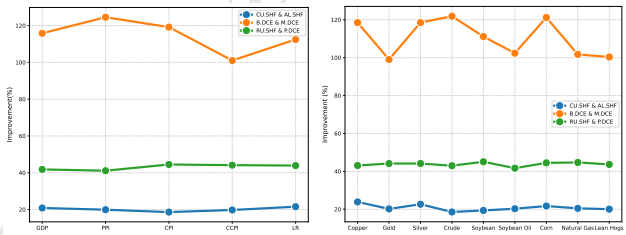


Figure 3: Trading signal analysis for CU.SHF & AL.SHF pair



(a) Macro **(b) US Commodity**

Figure 4: Performance improvement in annual return (%) when incorporating (a) Chinese macroeconomic indicators and (b) U.S. commodity futures.

Chinese Macroeconomic Indicators. To assess the predictive utility of domestic macroeconomic signals, we individually introduce five indicators: GDP, PPI, CPI, CCPI, and LPR. As shown in Figure 4a, all pairs exhibit consistent gains, confirming the relevance of macroeconomic context in financial modeling.

Among the three pairs, B.DCE & M.DCE shows the most pronounced benefit from macro features, with the greatest improvement observed from PPI, exceeding 120%. This highlights the strong connection between upstream production costs and chemical/energy-linked commodities. For RU.SHF & P.DCE, consistent gains are observed, with CPI and CCPI contributing slightly more than other macro indicators, suggesting that consumer price dynamics carry predictive signals for industrial consumer-related commodities. The CU.SHF & AL.SHF pair sees the highest improvements from LPR (approx. 22%) and GDP (approx. 21%), indicating that monetary policy and growth expectations have notable influence on base metals. Overall, the inclusion of macroeconomic variables enhances model performance across all sectors, with the most influential indicator varying by commodity type.

U.S. Commodity Futures. We next examine the effect of incorporating international market signals by introducing U.S.commodity futures. Figure 4b illustrates performance improvements for each pair across nine selected U.S. commodities.

Table 2: Performance Comparison Across All Methods and Pairs. The best results for CAR, AR, and SR are highlighted in bold.

Pair	Method	CAR (%)	AR (%)	SR	MDD	Total Holding Day
CU.SHF & AL.SHF	B&H	0.054	0.057	0.490	0.0511	342
	SA	0.013	0.014	-0.208	0.021	46
	DQN	10.865 ± 2.409	2.730 ± 0.224	**4.187 ± 0.345**	$(-8.08 \pm 0.15) \times 10^{-4}$	198 ± 54
	Double DQN	9.921 ± 4.806	2.525 ± 0.628	3.892 ± 0.929	$(\mathbf{-8.06 \pm 0.051}) \times 10^{-4}$	216 ± 76
	Ours	**13.352 ± 1.481**	**2.946 ± 0.1056**	2.408 ± 0.248	$(-8.10 \pm 0.07) \times 10^{-4}$	206 ± 25
B.DCE & M.DCE	B&H	0.0015	0.0016	-1.52	0.014	342
	SA	0.003	0.003	-3.993	0.002	97
	DQN	0.269 ± 0.136	0.242 ± 0.108	5.437 ± 1.968	$(\mathbf{-9.32 \pm 0.14}) \times 10^{-5}$	186 ± 71
	Double DQN	0.326 ± 0.113	0.291 ± 0.087	6.317 ± 1.559	$(-1.15 \pm 0.98) \times 10^{-4}$	253 ± 41
	Ours	**0.381 ± 0.065**	**0.335 ± 0.049**	**7.108 ± 0.826**	$(-1.15 \pm 0.99) \times 10^{-4}$	237 ± 25
RU.SHF & FU.SHF	B&H	0.016	0.016	-0.138	0.024	342
	SA	-0.005	-0.005	-2.430	0.016	73
	DQN	2.611 ± 0.852	1.327 ± 0.289	5.512 ± 1.084	$(-1.60 \pm 0.20) \times 10^{-4}$	180 ± 62
	Double DQN	2.831 ± 1.016	1.378 ± 0.345	5.723 ± 1.303	$(\mathbf{-1.59 \pm 0.0054}) \times 10^{-4}$	214 ± 69
	Ours	**3.920 ± 0.305**	**1.683 ± 0.064**	**6.931 ± 0.278**	$(-1.59 \pm 0.01) \times 10^{-4}$	216 ± 17
RU.SHF & PP.DCE	B&H	0.017	0.018	-0.0875	0.0150	342
	SA	-0.013	-0.013	-3.553	0.019	62
	DQN	1.575 ± 0.590	0.960 ± 0.290	5.922 ± 1.539	$(-2.00 \pm 0.16) \times 10^{-4}$	189 ± 69
	Double DQN	2.071 ± 0.338	1.173 ± 0.119	7.091 ± 0.694	$(-2.05 \pm 0.010) \times 10^{-4}$	234 ± 23
	Ours	**2.121 ± 0.249**	**1.193 ± 0.081**	**7.212 ± 0.480**	$(\mathbf{-1.59 \pm 0.01}) \times 10^{-4}$	209 ± 23
RU.SHF & P.DCE	B&H	0.022	0.023227	0.114272	0.024503	342
	SA	0.002	0.002	-1.128	0.013	100
	DQN	1.726 ± 0.432	1.037 ± 0.206	6.355 ± 1.112	$(-2.10 \pm 0.14) \times 10^{-4}$	213 ± 53
	Double DQN	0.743 ± 0.691	0.512 ± 0.396	3.476 ± 2.182	$(\mathbf{-2.10 \pm 0.016}) \times 10^{-4}$	176 ± 111
	Ours	**2.141 ± 0.368**	**1.195 ± 0.130**	**7.239 ± 0.742**	$(-2.11 \pm 0.02) \times 10^{-4}$	199 ± 32

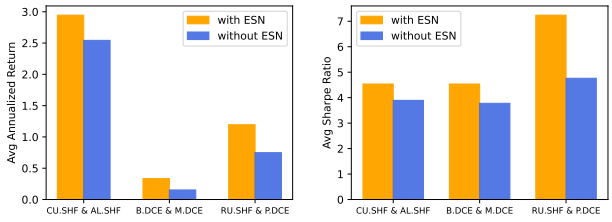
Table 3: Average Performance Summary Across All Pairs

Method	CAR	AR	Sharpe Ratio	MDD
B&H	0.0221	0.0232	-0.2282	0.0257
SA	0.0000	0.0002	-2.624	0.0140
DQN	3.4092	1.2592	5.4826	-0.0002
Double DQN	3.1784	1.1758	5.2998	-0.0002
Ours	4.3830	1.4704	6.1796	-0.0002

Table 4: Comparison of Macro vs. US Performance Metrics

Pair	Data Source	AR	Sharpe Ratio	MDD
CU.SHF & AL.SHF	Macro	2.3948	0.1632	-8.05×10^{-4}
	US	2.3609	0.2576	-8.15×10^{-4}
B.DCE & M.DCE	Macro	0.1414	0.0409	-9.30×10^{-5}
	US	0.0958	0.0169	-9.58×10^{-5}
RU.SHF & P.DCE	Macro	1.0905	0.0452	-2.097×10^{-4}
	US	0.9003	0.2664	-2.14×10^{-4}

All three pairs benefit from these external features, with B.DCE & M.DCE again exhibiting the most substantial gains. Improvements range from 100% to over 120%, particularly when incorporating Copper, Crude, and Corn futures—commodities that share demand/supply linkages or industrial use with domestic products.



(a) Annual return **(b) Sharp ratio**

Figure 5: Efectiveness of the ESN module

RU.SHF & P.DCE and CU.SHF & AL.SHF pairs also show steady enhancements, though the magnitudes are relatively lower, typically around 40–45% and 20–25%, respectively. These results underscore the informational value of global commodity markets in forecasting domestic price dynamics. The effectiveness of U.S. futures in boosting trading performance supports the integration of cross-border data into modern quantitative trading systems.

4.2.5 Effective of the dule module. To assess the contribution of the External Stream Network (ESN) module within our dual-stream framework, we conduct an ablation study on the same three representative commodity pairs introduced in Section 4.2.4: CU.SHF & AL.SHF, B.DCE & M.DCE, and RU.SHF & P.DCE.

Figure 5 compares the performance of the full model (with ESN, orange bars) against a reduced version without the ESN module

(blue bars), using two evaluation metrics: annualized return (Figure 5a and Sharpe ratio (Figure 5b).

The inclusion of the ESN consistently improves both profitability and risk-adjusted performance across all pairs. Notably, the RU.SHF & P.DCE pair exhibits the most significant improvement, with the annualized return increasing from approximately 0.75% to 1.25%, and the Sharpe ratio rising from 4.5 to over 7.0. The other two pairs show similar trends, though with more moderate gains. These results underscore the effectiveness of the ESN module in capturing exogenous signals, thereby enhancing the robustness and overall performance of the trading strategy.

5 Conlustion

In this paper, we propose a dual-stream deep reinforcement learning framework for pair trading in the Chinese commodity futures market, designed to capture both endogenous price dynamics and exogenous contextual signals. By integrating domestic macroeconomic indicators and U.S. commodity futures into a unified decision-making model, our approach addresses key challenges related to data heterogeneity and market uncertainty. Our framwork is a general framework in the sense that it can incorporate different sources of information and also not restricted to any particular market.

Through extensive empirical evaluation on a decade of real-world trading data, we demonstrate that the proposed method consistently achieves superior performance compared to baselines, both in terms of returns and Sharpe ratio, which highlights the potential of context-aware reinforcement learning strategies in financial applications. Future work will explore online learning extensions and the application of our framework to other global futures markets.

References

[1] Richard Bellman. 1957. A Markovian decision process. *Journal of mathematics and mechanics* (1957), 679–684.

[2] Eduardo Borensztein and Carmen M Reinhart. 1994. The macroeconomic determinants of commodity prices. *Staff Papers* 41, 2 (1994), 236–261.

[3] Andrew Brim. 2020. Deep Reinforcement Learning Pairs Trading with a Double Deep Q-Network. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. 0222–0227.

[4] Andrew Brim. 2020. Deep reinforcement learning pairs trading with a double deep Q-network. In *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 0222–0227.

[5] Victor Chang, Xiaowen Man, Qianwen Xu, and Ching-Hsien Hsu. 2021. Pairs trading on different portfolios based on machine learning. *Expert Systems* 38, 3 (2021), e12649.

[6] Yu-Ying Chen, Wei-Lun Chen, and Szu-Hao Huang. 2018. Developing arbitrage strategy in high-frequency pairs trading with filterbank CNN algorithm. In *2018 IEEE International Conference on Agents (ICA)*. IEEE, 113–116.

[7] Yin-Wong Cheung and Kon S Lai. 1995. Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business & Economic Statistics* 13, 3 (1995), 277–280.

[8] FengSheng Chien, Ka Yin Chau, Muhammad Sadiq, and Ching-Chi Hsu. 2022. The impact of economic and non-economic determinants on the natural resources commodity prices volatility in China. *Resources Policy* 78 (2022), 102863.

[9] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2015. Gated feedback recurrent neural networks. In *International conference on machine learning*. PMLR.

[10] Binh Do and Robert Faff. 2010. Does simple pairs trading still work? *Financial Analysts Journal* 66, 4 (2010), 83–95.

[11] Robert J Elliott, John Van Der Hoek*, and William P Malcolm. 2005. Pairs trading. *Quantitative Finance* 5, 3 (2005), 271–276.

[12] Saeid Fallahpour, Hasan Hakimian, Khalil Taheri, and Ehsan Ramezanifar. 2016. Pairs trading strategy optimization using the reinforcement learning method: a cointegration approach. *Soft Computing* 20 (2016), 5051–5066.

[13] Adrian Fernandez-Perez, Bart Frijns, Ivan Indriawan, and Yiuman Tse. 2020. Pairs trading of Chinese and international commodities. *Applied Economics* 52, 48 (2020), 5203–5217.

[14] Hung-Gay Fung, Wai K Leung, and Xiaoqing Eleanor Xu. 2003. Information flows between the US and China commodity futures trading. *Review of Quantitative Finance and Accounting* 21, 3 (2003), 267–285.

[15] Evan Gatev, William N Goetzmann, and K Geert Rouwenhorst. 2006. Pairs trading: Performance of a relative-value arbitrage rule. *The review of financial studies* 19, 3 (2006), 797–827.

[16] Eli Hadad, Sohail Hodarkar, Beakal Lemeneh, and Dennis Shasha. 2024. Machine Learning-Enhanced Pairs Trading. *Forecasting* 6, 2 (2024), 434–455.

[17] Weiguang Han, Boyi Zhang, Qianqian Xie, Min Peng, Yanzhao Lai, and Jimin Huang. 2023. Select and trade: Towards unified pair trading with hierarchical reinforcement learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 4123–4134.

[18] Dieter Hess, He Huang, and Alexandra Niessen. 2008. How do commodity futures respond to macroeconomic news? *Financial Markets and Portfolio Management* 22, 2 (2008), 127–146.

[19] Nicolas Huck. 2009. Pairs selection and outranking: An application to the S&P 100 index. *European Journal of Operational Research* 196, 2 (2009), 819–825.

[20] Qiang Ji and Ying Fan. 2016. How do China's oil markets affect other commodity markets both domestically and internationally? *Finance Research Letters* 19 (2016), 247–254.

[21] Sang-Ho Kim, Deog-Yeong Park, and Ki-Hoon Lee. 2022. Hybrid deep reinforcement learning for pairs trading. *Applied Sciences* 12, 3 (2022), 944.

[22] Taewook Kim and Ha Young Kim. 2019. Optimizing the Pairs-Trading Strategy Using Deep Reinforcement Learning with Trading and Stop-Loss Boundaries. *Complexity* 2019, 1 (2019), 3582516.

[23] Christopher Krauss. 2017. Statistical arbitrage pairs trading strategies: Review and outlook. *Journal of Economic Surveys* 31, 2 (2017), 513–545.

[24] Yan-Xia Lin, Michael McCrae, and Chandra Gulati. 2006. Loss protection in pairs trading through minimum profit bounds: A cointegration approach. *Advances in Decision Sciences* 2006 (2006).

[25] Qingfu Liu and Yunbi An. 2011. Information transmission in informationally linked markets: Evidence from US and Chinese commodity futures markets. *Journal of International Money and Finance* 30, 5 (2011), 778–795.

[26] Malik Magdon-Ismail and Amir F Atiya. 2004. Maximum drawdown. *Risk Magazine* 17, 10 (2004), 99–102.

[27] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. 2015. Human-level control through deep reinforcement learning. *nature* 518, 7540 (2015), 529–533.

[28] Masao Ogaki and Joon Y Park. 1997. A cointegration approach to estimating preference parameters. *Journal of Econometrics* 82, 1 (1997), 107–134.

[29] Hossein Rad, Rand Kwong Yew Low, and Robert Faff. 2016. The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance* 16, 10 (2016), 1541–1558.

[30] Hamed Rad, Rosmy K. Y. Low, and Robert Faff. 2016. The profitability of pairs trading strategies: distance, cointegration and copula methods. *Quantitative Finance* 16, 10 (2016), 1541–1558.

[31] Mr Shaun K Roache. 2012. *China's Impacton World Commodity Markets*. International Monetary Fund.

[32] David Ruppert and David S Matteson. 2011. *Statistics and data analysis for financial engineering*. Vol. 13. Springer.

[33] Simão Moraes Sarmento and Nuno Horta. 2020. Enhancing a pairs trading strategy with the application of machine learning. *Expert Systems with Applications* 158 (2020), 113490.

[34] William F Sharpe. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance* 19, 3 (1964), 425–442.

[35] R Todd Smith and Xun Xu. 2017. A good pair: alternative pairs-trading strategies. *Financial Markets and Portfolio Management* 31 (2017), 1–26.

[36] Guanglin Sun, Jianfeng Li, and Zezhong Shang. 2022. Return and volatility linkages between international energy markets and Chinese commodity market. *Technological Forecasting and Social Change* 179 (2022), 121642.

[37] Stephen J Taylor. 2008. *Modelling financial time series*. world scientific.

[38] Ganapathy Vidyamurthy. 2004. *Pairs Trading: quantitative methods and analysis*. John Wiley & Sons.

[39] Fenghua Wen, Zhen Liu, Zhifeng Dai, Shaoyi He, and Wenhua Liu. 2022. Multi-scale risk contagion among international oil market, Chinese commodity market and Chinese stock market: A MODWT-Vine quantile regression approach. *Energy Economics* 109 (2022), 105957.

[40] Fucui Xu and Shan Tan. 2021. Dynamic Portfolio Management Based on Pair Trading and Deep Reinforcement Learning. In *Proceedings of the 2020 3rd International Conference on Computational Intelligence and Intelligent Systems*.

[41] Baochen Yang, Yingjian Pu, and Yunpeng Su. 2020. The financialization of Chinese commodity markets. *Finance Research Letters* 34 (2020), 101438.

[42] Yurun Yang, Ahmet Goncu, and Athanasios Pantelous. 2017. Pairs trading with commodity futures: evidence from the Chinese market. *China Finance Review*

International 7, 3 (2017), 274–294.

[43] Libo Yin and Liyan Han. 2016. Macroeconomic impacts on commodity prices: China vs. the United States. *Quantitative Finance* 16, 3 (2016), 489–500.

[44] Lizi Zhang. 2021. Pair trading with machine learning strategy in China Stock Market. In *2021 2nd International Conference on Artificial Intelligence and Information Systems*. 1–6.

[45] Xiaoyu Zhang and Yongfu Liu. 2020. The dynamic impact of international agricultural commodity price fluctuation on Chinese agricultural commodity prices. *International Food and Agribusiness Management Review* 23, 3 (2020), 391–410.