

# Graph Learning for Foreign Exchange Rate Prediction and Statistical Arbitrage

Anonymous Authors

## Abstract

We propose a two-step graph learning approach for foreign exchange statistical arbitrages (FXSAs), addressing two key gaps in prior studies: the absence of graph-learning methods for foreign exchange rate prediction (FXRP) that leverage multi-currency and currency–interest rate relationships, and the disregard of the time lag between price observation and trade execution. In the first step, to capture complex multi-currency and currency–interest rate relationships, we formulate FXRP as an edge-level regression problem on a discrete-time spatiotemporal graph. This graph consists of currencies as nodes and exchanges as edges, with interest rates and foreign exchange rates serving as node and edge features, respectively. We then introduce a graph-learning method that leverages the spatiotemporal graph to address the FXRP problem. In the second step, we present a stochastic optimization problem to exploit FXSAs while accounting for the observation-execution time lag. To address this problem, we propose a graph-learning method that enforces constraints through projection and ReLU, maximizes risk-adjusted return by leveraging a graph with exchanges as nodes and influence relationships as edges, and utilizes the predictions from the FXRP method for the constraint parameters and node features. Moreover, we prove that our FXSA method satisfies empirical arbitrage constraints. The experimental results demonstrate that our FXRP method yields statistically significant improvements in mean squared error, and that the FXSA method achieves a 61.89% higher information ratio and a 45.51% higher Sortino ratio than a benchmark. Our approach provides a novel perspective on FXRP and FXSA within the context of graph learning.

## CCS Concepts

• Computing methodologies → Neural networks.

## Keywords

Foreign Exchange, Graph Learning, Statistical Arbitrage, Deep Learning

## 1 Introduction

We address the problem of predicting foreign exchange (FX) rates and exploiting FX statistical arbitrage (StatArb) opportunities. Consider a currency triplet: A, B, and C. We first exchange one unit of currency A for B, then convert the resulting amount of B to C, and finally exchange the resulting amount of C back into A. In theory, the final amount of A should be one [10], implying inherent interrelationships among FX rates. However, in practice, it sometimes exceeds one [2], allowing potential profits. In this paper, we predict FX rates by leveraging the interrelationships among currencies and exploit FX StatArbs (FXSAs) involving three or more currencies, which arise in practice.

Although the currencies in the triplet are theoretically interrelated—both among themselves and with interest rates (IRs)—no

prior work has studied a graph learning (GL) method to exploit these relationships in FX rate prediction (FXRP). As multiple currencies exist globally, combinatorially many triplets can be formed, each inducing its own theoretical interrelations. It is difficult to represent such structured relationships in a tabular format and to employ canonical neural architectures, such as multi-layer perceptrons (MLPs). Moreover, the interest rate parity (IRP) implies there is an interaction between the IRs of two countries and their FX rate. To effectively capture and leverage both triplet-based relationships and the IRP, we leverage GL for FXRP.

While the spot price and the actual trading price can differ due to the time lag between price observation and trade execution, such discrepancies have not been addressed in the existing literature (e.g., [25, 26, 34]) on FX arbitrage. This omission renders prior methods potentially inadmissible in practice and subject to look-ahead bias. Moreover, the price discrepancies introduced by the observation-execution time lag pose risks to investors’ trading, which have likewise been overlooked. To account for these discrepancies and the associated risks, we formulate a stochastic optimization problem for FXSA and propose a GL-based method to solve it.

To address multi-currency relationships, currency–IR interactions, and price discrepancies arising from execution time lags, we develop a two-step GL approach. In each step, we formulate a problem and set forth a corresponding GL method.

We formulate FXRP as an edge-level regression problem on a discrete-time spatiotemporal graph. We define the graph with currencies and exchange relations as nodes and edges, respectively, where IRs and FX rates serve as node and edge features, respectively. We train a simple graph neural network (GNN) on this graph. The GNN also uses currency-value features we derive from FX rates under the maximum likelihood estimation (MLE) principle.

To exploit FXSAs, we formulate a stochastic optimization problem that accounts for the time lag between price observation and trade execution. We then propose a GL method that generates FX trading quantities to address it. The GL method solves the problem by enforcing constraints via projection and ReLU, and maximizing risk-adjusted return, while utilizing the predictions from the FXRP method for the constraint parameters and node features. We prove our FXSA method guarantees the satisfaction of the constraints. We construct a graph with exchanges as nodes and influence relationships—derived from the projection matrix—as edges, and train a GNN on the graph.

We demonstrate the effectiveness of our method on ten of the most traded currencies. Our FXSA method outperforms a benchmark across key risk-adjusted return metrics, exhibiting a steady upward trend in cumulative profits while reducing risk. Our FXRP results show that incorporating the FX graph with IRs leads to statistically significant improvements over non-graph-based methods.

The prominent contributions of this paper are as follows.

- To the best of our knowledge, we are the first to propose a spatiotemporal GL problem for FXRP using FX rates and government bond IRs, and present a dedicated GL method.
- To the best of our knowledge, we are the first to formulate FXSA as a stochastic optimization problem, propose a corresponding GL method that generates trading strategies to exploit arbitrages among three or more currencies, and prove that our FXSA method satisfies the empirical arbitrage constraints, thereby reducing risk.
- We demonstrate the effectiveness of our approach: our FXSA method yields a 61.89% higher information ratio and a 45.51% higher Sortino ratio than a benchmark, and the FXRP method shows statistically significant improvement over non-graph models.

## 2 Related Work

Existing learning-based studies overlook the interrelationships among FX rates. Prior econometric studies on FXRP commonly use models such as VAR [9] and GARCH [20]. However, these methods require underlying statistical assumptions and often fail to capture nonlinear dynamics [3]. To address these limitations, learning-based methods continue to gain traction. Deep learning models (e.g., Transformer [31], LSTM [16], CNN [15, 31]) and machine learning algorithms (e.g., gradient boosting [36], SVM [11, 23]) are being actively explored. Researchers apply such methods to diverse currency pairs (e.g., GBPJPY [23], USDCNY [31], USDCAD [16]), with major pairs like USDEUR [1, 16, 22, 23], USDJPY [22, 23], USDGBP [16, 22, 23], USDCHF [16] receiving the most attention [5]. However, these studies overlook the relationships among currency triplets—discussed in Section 1—despite their strong theoretical basis in economics [2, 10].

Prior studies do not consider the macroeconomic factors of currencies comprehensively. Theories such as the IRP and purchasing power parity delineate how an FX rate for a currency pair relates to macroeconomic factors [27]. To utilize these theories, prior studies incorporate macroeconomic features of a target FX pair, such as IRs [4, 8, 21, 35], inflation rates [21, 35], and monetary policies [8, 21]. However, the macroeconomic parity valuation of a currency can vary depending on the counterpart currency, due to market frictions (e.g., creditworthiness [30], banking regulation [12]). Thus, accurate FXRP requires comprehensively leveraging macroeconomic factors of both the target pair and other currencies—an aspect overlooked by prior learning-based studies.

The time lag between price observation and execution is inevitable in practice, but prior studies overlook it and its associated risks. Several studies show the existence of FX arbitrage opportunities in real markets [2, 13, 18]. To exploit them, prior approaches employ integer programming [25], the simulated bifurcation algorithm [26], and a GNN [34] to solve arbitrage optimization problems. However, these studies disregard the time lag between price observation and execution, rendering them susceptible to look-ahead bias and the risk of price movements between the observation and execution time points. To be practically applicable, such methods need to provide very accurate FXRP or address the risks stemming from the time lag. In addition, [34] is prone to violating flow-in-flow-out constraints due to its use of a relaxed loss with a squared

penalty to satisfy the constraints, and it uses synthetic data. Unlike the questionable assumptions of [34], our problem formulation accounts for the time lag and FX rate stochasticity. Moreover, our method provably satisfies all constraints, utilizes predictions from our FXRP method to address the time lag and stochasticity, and is evaluated on real market data.

To address the limitations of the prior studies, we propose a two-stage GL approach. The first formulates FXRP as a GL problem and presents a GL method, capturing interrelationships among FX rates and comprehensive interest-rate-based valuations. Building on these predictions, the second formulates FXSA as a stochastic optimization problem that addresses the observation–execution time lag and its associated risks. It also proposes a GL method that provably satisfies the empirical arbitrage constraints.

## 3 Notation and Preliminaries

**Notation.** We map all time points on trading dates to  $\mathcal{T} = (0, \infty)$  while preserving their temporal order, such that each natural number corresponds to the end time point (24:00) of a trading date. We use this end time to represent the trading date itself. That is, for a time point  $\tau \in \mathcal{T}$ , the ceiling  $\lceil \tau \rceil \in \mathbb{N}$  denotes the corresponding trading date. We let  $t \in \mathbb{N} \subset \mathcal{T}$  and  $\tau \in \mathcal{T}$  denote a trading date and a time point, respectively.

We consider a single agent investing in the FX market. Let  $C$  be the set of investable currencies, where  $o \in C$  denotes the agent’s home currency. Let  $Y_{\tau, \Delta\tau, i} \in \mathbb{R}$  and  $Y_{\tau, \Delta\tau, i}^f \in \mathbb{R}$  denote the government bond and risk-free IRs of currency  $i \in C$ , respectively, quoted at time  $\tau \in \mathcal{T}$  for the period from  $\tau$  to  $\tau + \Delta\tau$ . For  $\tau \in \mathcal{T}$  and  $i, j \in C$ , let  $X_{\tau ij} > 0$  be a random variable denoting the FX rate from  $i$  to  $j$  at  $\tau$ , defined as the amount of  $j$  per unit of  $i$ . Let  $E_{\tau} \subset C \times C$  be the set of currency pairs that can be directly exchanged at  $\tau$ . For each  $(i, j) \in E_{\tau}$ , let  $w_{\tau ij} \geq 0$  denote the amount, in units of currency  $o$ , that a trading strategy allocates to exchange from  $i$  to  $j$ .

We use the tilde symbol to denote the realization of a random variable, and the hat symbol to denote the agent’s estimate of its expectation. For any vectors  $\mathbf{x}_i \in \mathbb{R}^p$  with  $i \in I$ , where  $I$  is a finite index set, we denote by  $[\mathbf{x}_i]_{i \in I} \in \mathbb{R}^{|I| \times p}$  the matrix formed by stacking  $\mathbf{x}_i$  row-wise. For any  $a, b \in \mathbb{R}$ , we define  $[a]^+ = \max\{a, 0\}$ . We let  $[a : b] = [a, b] \cap \mathbb{Z}$  and use similar notation for open or half-open intervals.

**Preliminaries.** An FX direct arbitrage refers to a riskless cash flow arising from a violation of

$$X_{\tau ij} X_{\tau ji} = 1 \quad (1)$$

for  $\tau \in \mathcal{T}$  and  $i, j \in C$ . If  $X_{\tau ij} X_{\tau ji} > 1$  is known in advance, the agent can earn a profit of  $X_{\tau ij} X_{\tau ji} - 1$  by converting one unit of currency  $i$  into  $j$  and immediately converting it back to  $i$  at  $\tau$ . Throughout this paper, we assume (1) holds to focus on more intriguing arbitrages involving more than two currencies. It is easy to incorporate such two-currency arbitrages in what follows.

An FX triangular arbitrage refers to a riskless cash flow that arises from the violation of

$$X_{\tau ij} = X_{\tau ik} X_{\tau kj} \quad (2)$$

for  $\tau \in \mathcal{T}$  and  $i, j, k \in C$ . If both (1) and  $X_{\tau ij} < X_{\tau ik} X_{\tau kj}$  hold and the agent knows this in advance, then, since  $X_{\tau ik} X_{\tau kj} X_{\tau ji} > 1$ , the agent can earn a profit of  $X_{\tau ik} X_{\tau kj} X_{\tau ji} - 1$  by sequentially

exchanging from  $i$  to  $k$ , from  $k$  to  $j$ , and from  $j$  to  $i$ . There is evidence that (2) is violated in practice [2], although it should hold in theory [10], implying underlying multi-currency relationships.

Extending FX triangular arbitrage to cases involving more than three currencies, we consider an LP problem (3)–(6)—adapted from [25]. For  $\tau'', \tau', \tau \in \mathcal{T}$  with  $\tau'' < \tau' < \tau$ , suppose the agent observes  $[\tilde{X}_{\tau''ij}]_{(i,j) \in E_{\tau''}}$  at time  $\tau''$  and knows, in advance at  $\tau'$ , that  $\tilde{X}_{\tau ij} = \tilde{X}_{\tau''ij}$  for all  $(i, j) \in E_{\tau} \cap E_{\tau''}$ . Then, the agent can estimate the expectation of  $X_{\tau ij}$  as  $\hat{X}_{\tau ij} = \tilde{X}_{\tau''ij}$ , solve the LP problem (3)–(6), and execute the solution at  $\tau$  to maximize the profit from FX arbitrages. In (3), the agent maximizes the profit defined as the difference between total cash inflows into  $o$  and cash outflows from it. In (4), the agent does not retain any nonzero amount in currencies other than  $o$  to avoid the risk associated with holding a currency after trading, which implies the net cash flow for each  $i \in C \setminus \{o\}$  is zero. The first and second terms in (4) represent the total inflows to  $i$  and the total outflows from  $i$ , respectively. Since  $w_{\tau ij}$  is in units of  $o$ , the second term converts it to  $\hat{X}_{\tau oi} w_{\tau ij}$  in units of  $i$ . In the first term, the agent exchanges  $\hat{X}_{\tau oj} w_{\tau ji}$ , in units of  $j$ , for  $\hat{X}_{\tau oj} \hat{X}_{\tau ji} w_{\tau ji}$ , in units of  $i$ . Equation (5) imposes the unit trading constraint. If a different amount is traded, the solution can be scaled proportionally.

$$\max_{w_{\tau ij}: (i,j) \in E_{\tau} \cap E_{\tau''}} \sum_{j:(j,o) \in E_{\tau} \cap E_{\tau''}} w_{\tau jo} - \sum_{j:(o,j) \in E_{\tau} \cap E_{\tau''}} w_{\tau oj} \quad (3)$$

s.t.

$$\sum_{j:(j,i) \in E_{\tau} \cap E_{\tau''}} \hat{X}_{\tau oj} \hat{X}_{\tau ji} w_{\tau ji} - \sum_{j:(i,j) \in E_{\tau} \cap E_{\tau''}} \hat{X}_{\tau oi} w_{\tau ij} = 0, \forall i \in C \setminus \{o\}, \quad (4)$$

$$\sum_{(i,j) \in E_{\tau} \cap E_{\tau''}} w_{\tau ij} = 1, \quad (5)$$

$$w_{\tau ij} \geq 0, \forall (i, j) \in E_{\tau} \cap E_{\tau''}. \quad (6)$$

Since the assumption  $\tilde{X}_{\tau ij} = \tilde{X}_{\tau''ij}$  in (3)–(6) does not hold in practice, we need to predict  $X_{\tau ij}$ . To that end, we utilize IRs as input features based on the following IRP theories. Under the arbitrage-free assumption and ideal conditions [12], the covered IRP implies that for  $\tau \in \mathcal{T}$  and  $i, j \in C$ ,  $(1 + Y_{\tau, \Delta\tau, i}^f) F_{\tau, \Delta\tau, i, j} = (1 + Y_{\tau, \Delta\tau, j}^f) X_{\tau ij}$ , where  $F_{\tau, \Delta\tau, i, j}$  is the forward FX rate at  $\tau$  from  $i$  to  $j$  with maturity at  $\tau + \Delta\tau$ . If  $F_{\tau, \Delta\tau, i, j}$  is replaced by the conditional expectation of  $X_{\tau ij}$ , then the covered IRP identity becomes the uncovered IRP identity. Both versions of IRP imply FX rates are theoretically related to IRs.

## 4 Proposed Method

To determine daily FXSA trade lists, we propose a two-step GL approach: one for FXRP and the other one for FXSA. The first yields FXRP model  $f_P(\cdot; \theta_t^P)$ , and the second yields FXSA model  $f_S(\cdot; \theta_t^S)$  for  $t \in [t_1 : \infty)$ , where  $t_1 \in \mathbb{N}$  is the date when the agent starts trading. Here,  $\theta_t^P$  and  $\theta_t^S$  are trainable parameters. For each  $t \in [t_1 : \infty)$ ,  $f_P(\cdot; \theta_t^P)$  predicts  $\hat{X}_{t ij}$  for  $X_{t ij}$  using available information up to time  $t - 1$ . Then,  $f_S(\cdot; \theta_t^S)$  determines trading quantities  $w_{t ij}$  based on the same information and the predictions  $\hat{X}_{t ij}$  from  $f_P$  by time  $t - t_{\text{exec}}$ , where  $t_{\text{exec}} \in (0, 1)$  denotes the time required to execute FX trades at  $t$ . The agent simultaneously

executes all trades  $w_{t ij}$  at  $t$ , holds the resulting currencies until  $t + 1$ , and converts all holdings to the home currency  $o$  at  $t + 1$ .

The FXRP and FXSA models are trained on certain trading dates and used until the next training date. Given hyperparameters  $t_1$  and  $n_{\text{fit}} \in \mathbb{N}$ , we define a sequence  $\{t_k\}_{k=0}^{n_{\text{fit}}+1} \subset \mathbb{N} \cup \{\infty\}$  such that  $t_0 = 1$ ,  $t_k < t_{k+1}$  for all  $k \in [0 : n_{\text{fit}}]$ , and  $t_{n_{\text{fit}}+1} = \infty$ . Then, we define the testing periods  $\mathcal{T}_k^{\text{te}} = [t_k : t_{k+1})$  for all  $k \in [1 : n_{\text{fit}}]$ , forming a partition of  $[1 : n_{\text{fit}}+1)$ . For each  $k \in [1 : n_{\text{fit}}]$ , during the period  $(t_k - 1, t_k - t_{\text{exec}})$ , we train  $f_P(\cdot; \theta_k^P)$  and  $f_S(\cdot; \theta_k^S)$  using the training and validation periods  $(\mathcal{T}_k^{P, \text{tr}}, \mathcal{T}_k^{P, \text{va}})$  and  $(\mathcal{T}_k^{S, \text{tr}}, \mathcal{T}_k^{S, \text{va}})$ , respectively, where  $\mathcal{T}_k^{P, \text{tr}} \cup \mathcal{T}_k^{P, \text{va}} = \mathcal{T}_k^{S, \text{tr}} \cup \mathcal{T}_k^{S, \text{va}} = [t_0 : t_k)$  and  $\mathcal{T}_k^{P, \text{tr}} \cap \mathcal{T}_k^{P, \text{va}} = \mathcal{T}_k^{S, \text{tr}} \cap \mathcal{T}_k^{S, \text{va}} = \emptyset$ . We then fix  $\theta_t^P = \theta_k^P$  and  $\theta_t^S = \theta_k^S$  for  $t \in \mathcal{T}_k^{\text{te}}$ .

### 4.1 Foreign Exchange Rate Prediction

We formulate FXRP as an edge-level regression problem on a discrete-time spatiotemporal graph. For each  $t \in \mathcal{T}$ , we define a graph  $\mathcal{G}_t^P = (C, E_t, [Y_{t, i, i}]_{i \in C}, [X_{t ij}]_{(i, j) \in E_t})$ , whose components represent the node set, edge set, node features, and edge features, respectively. We define  $\mathcal{G}_{1:t}^P$  as a collection  $\{\mathcal{G}_{t'}^P : t' \in [1 : t]\}$ . We define  $f_P(\mathcal{G}_{1:t-1}^P; \theta_t^P) = h_{PO}(g_P(h_{PI}(\mathcal{G}_{1:t-1}^P); \theta_t^P)) = [\hat{X}_{t ij}]_{(i, j) \in \mathcal{U}_t}$  as the prediction function. Here,  $h_{PO}$  and  $h_{PI}$  are output-scaling and feature-engineering functions, respectively, and  $g_P$  is a GNN with parameters  $\theta_t^P$ . We formulate the FXRP problem as training  $f_P$  to predict  $\tilde{X}_{t ij}$  through the following loss:

$$\min_{\theta_k^P} \sum_{t \in \mathcal{T}_k^{P, \text{tr}}} d_P \left( g_P(h_{PI}(\mathcal{G}_{1:t-1}^P); \theta_k^P), h_{PO}^{-1}([\tilde{X}_{t ij}]_{(i, j) \in \mathcal{U}_t}) \right) \quad (7)$$

for each  $k \in [1 : n_{\text{fit}}]$ , where  $d_P$  is a metric, and  $\mathcal{U}_t \subset E_t$  is the set of predicted edges. We select the best among  $n_{\text{hyper}}$  hyperparameter configurations via grid search, based on performance on  $\mathcal{T}_k^{P, \text{va}}$ . In our formulation, GNNs can capture multi-currency and currency-IR relationships through the graphs incorporating both FX rates and IRs.

To extract temporal information, we define  $h_{PI}(\mathcal{G}_{1:t}^P) = \mathcal{G}_{1:t}^{PF} = (C, L_t \cap L_{t-1}, [\mathbf{c}_{ti}]_{i \in C}, [\mathbf{x}_{t ij}]_{(i, j) \in L_t \cap L_{t-1}})$ . Here, we define  $L_t = \{(i, j) \in E_t : (j, i) \in E_t\}$ . As (1) is assumed, we consider only reciprocal edges, i.e.,  $L_t$  instead of  $E_t$ . We define  $\mathcal{T}(i, j, t, t') = \{t'' \in [t - t' + 1 : t] : (i, j) \in L_{t''} \cap L_{t''-1}\}$  as the set of dates  $t''$  in  $[t - t' + 1 : t]$  such that both exchanges  $(i, j)$  and  $(j, i)$  are directly tradable on both  $t''$  and  $t'' - 1$ . We define  $\mathcal{T}^R \subset \mathbb{N}$  as a set of look-back windows, which are hyperparameters. We define  $\mathbf{x}_{t ij} = [x_{t ij t'}]_{t' \in \mathcal{T}^R}$  where  $x_{t ij t'} = \frac{1}{|\mathcal{T}(i, j, t, t')|} \sum_{t'' \in \mathcal{T}(i, j, t, t')} \log(\tilde{X}_{t'' ij} / \tilde{X}_{t''-1, i, j})$ . Here,  $x_{t ij t'}$  denotes the average temporal log difference,  $\log(\tilde{X}_{t'' ij} / \tilde{X}_{t''-1, i, j})$ , computed over the past  $t'$  dates (i.e.,  $t'' \in [t - t' + 1 : t]$ ) during which both  $(i, j)$  and  $(j, i)$  are directly tradable. The vector  $\mathbf{x}_{t ij}$  collects these averages across different look-back windows  $t' \in \mathcal{T}^R$ . We define  $[\mathbf{c}_{ti}]_{i \in C} = [\mathbf{y}_{ti}; \mathbf{v}_{ti}] \in \mathbb{R}^{2|\mathcal{T}^R|}$  where  $\mathbf{y}_{ti} = [y_{t i t'}]_{t' \in \mathcal{T}^R}$  and  $\mathbf{v}_{ti} = [v_{t i t'}]_{t' \in \mathcal{T}^R}$ . Here, we define  $y_{t i t'} = \frac{1}{t'} \sum_{t''=t-t'+1}^t \log\left(\frac{1 + \tilde{Y}_{t'', 1, i}}{1 + \tilde{Y}_{t''-1, 1, i}}\right)$  and  $v_{t i t'} = \frac{1}{t'} \sum_{t''=t-t'+1}^t \log(\tilde{V}_{t'', i} / \tilde{V}_{t''-1, i})$  as the average temporal log differences of the interest rate and the currency value, respectively, over the past  $t'$  dates, where

$\tilde{V}_{ti}$  is discussed next. These averages across different look-back windows  $t' \in \mathcal{T}^R$  are collected in the vectors  $\mathbf{y}_{ti}$  and  $\mathbf{v}_{ti}$ .

We define  $\tilde{V}_{ti}$  as the least-squares solution to (8) and (9):

$$\log \tilde{V}_{ti} - \log \tilde{V}_{tj} = \log \tilde{X}_{tij}, \forall (i, j) \in L_t : i < j, \quad (8)$$

$$\frac{1}{|C|} \sum_{i \in C} \log \tilde{V}_{ti} = 0, \quad (9)$$

which is MLE in the market. Let  $V_{ti}$  denote the value of currency  $i$  in units of an imaginary currency  $o^*$ . From (2), for each  $(i, j) \in L_t$ , the exchange rate is modeled as  $X_{tij} = \exp(\alpha_{tij}) \cdot V_{ti}/V_{tj}$ , or equivalently,  $\log X_{tij} = \log V_{ti} - \log V_{tj} + \alpha_{tij}$ . Since  $\alpha_{tij} \neq 0$  implies arbitrage, and such opportunities are expected to be exploited [17, 32], we assume  $\mathbb{E}\alpha_{tij} = 0$ . For analytical tractability, we assume  $\alpha_{tij} \sim \mathcal{N}(0, \sigma_\alpha^2)$ . MLE maximizes the log-likelihood of  $\alpha_{tij}$ :  $\max [-\frac{L_t l}{2} \log(2\pi\sigma_\alpha^2) - \frac{1}{2\sigma_\alpha^2} \sum_{(i,j) \in L_t} (\log \tilde{V}_{ti} - \log \tilde{V}_{tj} - \log \tilde{X}_{tij})^2]$ . Its maximizer is the least-squares solution to (8), where the unknowns are  $\log \tilde{V}_{ti}$  and  $\log \tilde{V}_{tj}$ . As (1) is assumed, we restrict (8) to cases where  $i < j$ . Since the coefficient matrix of (8) is a complete incidence matrix, it has rank  $|C| - 1$  [7], and thus we add (9). Then, (8) and (9) have rank  $|C|$ <sup>1</sup>, implying a unique least-squares solution always exists.

We define  $h_{PO}^{-1}([X'_{tij}]_{(i,j) \in \mathcal{U}_t}) = [\log(X'_{tij}/\tilde{X}_{t-1,ij})]_{(i,j) \in \mathcal{U}_t}$  and define  $h_{PO}$  as the inverse of  $h_{PO}^{-1}$ . We set  $\mathcal{U}_t = L_t \cap L_{t-1} \cap L_{t-2}$  since  $h_{PI}(\mathcal{G}_{1:t-1}^P)$  uses temporal differences between  $t-1$  and  $t-2$ .

We use a simple GNN  $g_P$  that handles both node and edge features. Given a graph  $\mathcal{G}^0 = (N, E, [\mathbf{n}_i^0]_{i \in N}, [\mathbf{e}_{ij}^0]_{(i,j) \in E})$ , we define  $g_{C,l}(\mathcal{G}^{l-1}) = \mathcal{G}^l = (N, E, [\mathbf{n}_i^l]_{i \in N}, [\mathbf{e}_{ij}^l]_{(i,j) \in E})$  for each  $l \in [1 : L]$ , where

$$\mathbf{n}_i^l = \frac{1}{|N(i)|} \sum_{j \in N(i)} \text{SLP}_{N,l}([\mathbf{n}_i^{l-1}; \mathbf{e}_{ji}^{l-1}; \mathbf{n}_j^{l-1}]), \forall i \in N, \quad (10)$$

$$\mathbf{e}_{ij}^l = \text{SLP}_{E,l}([\mathbf{n}_i^l; \mathbf{e}_{ij}^{l-1}; \mathbf{n}_j^l]), \forall (i, j) \in E, \quad (11)$$

$N(i) = \{j : (j, i) \in E\}$ , and SLP denotes a single-layer feedforward network. Let  $g_{CN,l}(\mathcal{G}^0) = [\mathbf{n}_i^l]_{i \in N}$  and  $g_{CE,l}(\mathcal{G}^0) = [\mathbf{e}_{ij}^l]_{(i,j) \in E}$ . Then, we define  $g_P$  with  $L$  layers as  $g_P(h_{PI}(\mathcal{G}_{1:t-1}^P)) = g_P(\mathcal{G}_{t-1}^{PF}) = \text{SLP}_H(g_{CE,L}(g_{CS}(\mathcal{G}_{t-1}^{PF})))$ , where  $\text{SLP}_H$  has no activation function, and  $g_{CS}$  scales the node and edge features of  $\mathcal{G}_{t-1}^{PF}$ .

By (10) and (11),  $g_P(h_{PI}(\mathcal{G}_{1:t-1}^P))$  can capture complex multi-currency-IR relationships. In (10), each currency aggregates information about the IRs of its neighbors and the FX rates connected to it. Using such information from two currencies, the embedding of their FX rate is updated in (11). Over multiple layers of (10) and (11),  $g_P$  extracts complex relationships between FX rates and IRs.

Since  $f_S$  uses the prediction  $[\tilde{X}_{tij}]_{(i,j) \in \mathcal{U}_t}$ , we impose an additional condition on  $\mathcal{T}_k^{P,va}$  such that  $\bigcup_{k=0}^{n_{\text{fit}}} \mathcal{T}_k^{P,va} = [t_0 : t_1]$ , where  $n_{\text{sy}} \in [1 : n_{\text{fit}}]$  is a hyperparameter. Then, for  $t \in [t_0, t_1]$ , we let  $k^*(t) = \min\{k' \in [1 : n_{\text{fit}}] : t \in \mathcal{T}_{k'}^{P,va}\}$  and define  $[\tilde{X}_{tij}]_{(i,j) \in \mathcal{U}_t} = f_P(\mathcal{G}_{1:t-1}^P; \theta_{k^*(t)}^P)$ . This ensures that for all  $k \in [n_{\text{sy}} : n_{\text{fit}}]$ ,  $f_S$  can be trained only on predictions made outside the training period of  $f_P$ . Without this condition,  $f_S$  could be trained

<sup>1</sup>Since the left-hand side of (8) is zero when  $\log V_{ti} = 1/|C|$  for all  $i \in C$ , the coefficient vector of (9) is orthogonal to the row space of (8), and thus linearly independent of its rows.

on  $f_P$ 's own training predictions for  $t \in [t_0 : t_1]$ , i.e.,  $[t_0 : t_1] \setminus \bigcup_{k=1}^{n_{\text{fit}}} \mathcal{T}_k^{P,va} \neq \emptyset$ . We evaluate  $f_P$  on  $\mathcal{T}_k^{\text{te}}$  for all  $k \in [1 : n_{\text{fit}}]$ , and  $f_S$  for  $k \in [n_{\text{sy}} : n_{\text{fit}}]$ .

## 4.2 Foreign Exchange Statistical Arbitrage

**Problem Formulation.** We formulate FXSA as the following stochastic optimization problem: for each  $t \in \mathbb{N}$ ,

$$\max_{w_{tij} : (i,j) \in \mathcal{U}'_t} \frac{\mathbb{E}G_t}{\sqrt{\text{Var}(G_t)}} \quad \text{s.t.} \quad (12)$$

$$G_t = \sum_{i \in C} \frac{1 + Y_{t,1,i}^f}{1 + Y_{t,1,o}^f} X_{t+1,i,o} H_{ti}, \quad (13)$$

$$H_{ti} = \sum_{j : (j,i) \in \mathcal{U}'_t} X_{tji} \tilde{X}_{toj} w_{tji} - \sum_{j : (i,j) \in \mathcal{U}'_t} \tilde{X}_{toi} w_{tij}, \forall i \in C, \quad (14)$$

$$\mathbb{E}H_{ti} = 0, \forall i \in C \setminus \{o\}, \quad (15)$$

$$\sum_{(i,j) \in \mathcal{U}'_t} w_{tij} = 1, \quad (16)$$

$$w_{tij} w_{tji} = 0, \forall (i, j) \in \mathcal{U}'_t, \quad (17)$$

$$w_{tij} \geq 0, \forall (i, j) \in \mathcal{U}'_t. \quad (18)$$

Compared to [25, 26, 34], our problem includes the random variables  $X_{tji}, X_{t+1,i,o}$  in (13) and (14) to account for the time lag between observation at  $t-1$  and execution at  $t$ .

Deterministic decision variable  $w_{tij}$  is the amount (in units of  $o$ ) the agent converts from currency  $i$  to  $j$ . We define  $H_{ti} \in \mathbb{R}$ , for  $i \in C \setminus \{o\}$ , as the agent's holding in currency  $i$  from  $t$  to  $t+1$ ;  $H_{to} \in \mathbb{R}$  as the instantaneous profit or loss in  $o$  at  $t$ ; and  $G_t \in \mathbb{R}$  as the time- $t$  present-value gain from trading, in units of  $o$ . Since (13) and (14) involve  $\tilde{X}_{toi}$  and  $X_{t+1,i,o}$ , we restrict the link set to  $\mathcal{U}'_t \subset \{(i, j) \in \mathcal{U}_t : \{(o, i), (o, j)\} \subset \mathcal{U}_t \cup \{(o, o)\}, \{(i, o), (j, o)\} \subset E_t \cup \{(o, o)\}\}$ . As (1) is assumed, we remove the direct arbitrage trades—where both  $w_{tij}$  and  $w_{tji}$  are positive—by imposing (17).

In (12), our problem maximizes the information ratio as defined in [24], assuming a zero benchmark expected return. Since FX rates can change between  $t$  and  $t+1$ , i.e.,  $X_{tij} \neq X_{t+1,ij}$ , we account for the associated risk by maximizing the risk-adjusted return measured by the information ratio. Because we pursue a strategy that holds no position between  $t$  and  $t+1$ , i.e.,  $H_{ti} = 0, \forall i \in C \setminus \{o\}$ , as required for deterministic arbitrage, we assume  $\mathbb{E}[\sum_{i \in C \setminus \{o\}} H_{ti} X_{tio} Y_{t,1,o}^f] = 0$ . That is, we assume that the expected return of the benchmark that invests  $H_{ti}$  in the risk-free asset between  $t$  and  $t+1$  is zero. Since  $H_{to}$  is an instantaneous profit or loss at  $t$ , it can be invested through other strategies between  $t$  and  $t+1$ .

The agent trades at  $t$ , holds the resulting currencies  $\{H_{ti} : i \in C \setminus \{o\}\}$  until  $t+1$ , converts all holdings back to  $o$  at  $t+1$ , and then to their time- $t$  present values, as represented by (13). We assume that, for each  $t \in \mathbb{N}$ , the agent has unlimited borrowing and lending capacity at the rate  $Y_{t,1,i}^f$  in any currency  $i \in C \setminus \{o\}$ . Thus, during the holding period  $(t, t+1)$ , interest accrues, and each holding position becomes  $(1 + Y_{t,1,i}^f)H_{ti}$ , in units of  $i$ , at  $t+1$ . After converting all holdings to the home currency  $o$  and then to their time- $t$  values, the gain  $G_t$  from the trading strategy is given by (13). For notational convenience, we define  $X_{t,o,o} = \tilde{X}_{t,o,o} = \tilde{X}_{t,o,o} = 1$  for all  $t \in \mathbb{N}$ . Then, (13) and (14) can also be applied to  $H_{to}$ .

In (14) and (15), for  $i \in C \setminus \{o\}$ , the holding position  $H_{ti}$  is the net cash flow, and its expectation is constrained to be zero. In (14), the first and second terms denote total cash inflows and outflows, respectively. Since  $w_{tij}$  is in units of  $o$ , it is converted to  $i$  as  $\widehat{X}_{toi}w_{tij}$ , forming the second term. For the first term, the agent exchanges  $\widehat{X}_{toj}w_{tji}$  in currency  $j$  for  $i$ , resulting in  $X_{tji}\widehat{X}_{toj}w_{tji}$ . When converting from  $j$  to  $i$ , the amount in  $j$  should be decided by the agent before the exchange, while the resulting amount in  $i$  is determined by the market at execution. Thus,  $\widehat{X}_{toj}$  is an agent's estimate, while  $X_{tji}$  is a random variable. Due to the stochasticity of  $X_{tji}$ , we use (15) as a surrogate constraint to reduce the risk of FX rate changes between  $t$  and  $t+1$ , while we ideally pursue a strategy satisfying  $H_{ti} = 0, \forall i \in C \setminus \{o\}$ .

**Graph Learning.** To solve (12)–(18), we train  $f_S$  to learn an optimal strategy from realized samples via an empirical formulation. We define a GL model  $f_S(\mathcal{G}_{1:t-1}^P, [\widehat{X}_{tij}]_{(i,j) \in \mathcal{U}_t^S}; \theta_t^S) = h_{SO}(g_S(h_{SI}(\mathcal{G}_{1:t-1}^P, [\widehat{X}_{tij}]_{(i,j) \in \mathcal{U}_t^S}; \theta_t^S))) = [w_{tij}]_{(i,j) \in \mathcal{U}_t^S}$  and train it by minimizing the empirical objective corresponding to (12)–(14). Here,  $h_{SI}$  and  $g_S$  denote a feature-mapping function and GNN, respectively. The post-processing function  $h_{SO}$  ensures that the output satisfies the empirical constraints corresponding to (15)–(18). To satisfy (1), we redefine  $\widehat{X}_{tij}$  as the geometric average of the predictions from  $f_P$ :  $\widehat{X}_{tij}$  and  $1/\widehat{X}_{tji}$ .

To achieve (12)–(14), we train multiple GL models and select the best models. For each  $k \in [1 : n_{\text{fit}}]$ , we train  $f_S$  with loss

$$\min_{\theta_k^S} \mathbb{E}_{\mathcal{B}} \left[ -\frac{\hat{\mu}_G}{\hat{\sigma}_G^2} \cdot \mathbb{I}_{\{\hat{\mu}_G > 0\}} - \hat{\mu}_G \cdot \mathbb{I}_{\{\hat{\mu}_G \leq 0\}} \right] \quad \text{s.t.} \quad (19)$$

$$\hat{\mu}_G = \frac{1}{|\mathcal{B}|} \sum_{t \in \mathcal{B}} \widetilde{G}_t, \quad \hat{\sigma}_G^2 = \frac{1}{|\mathcal{B}| - 1} \sum_{t \in \mathcal{B}} (\widetilde{G}_t - \hat{\mu}_G)^2, \quad (20)$$

$$\widetilde{G}_t = \sum_{i \in C} \frac{1 + \widetilde{Y}_{t,1,i}}{1 + \widetilde{Y}_{t,1,o}} \widetilde{X}_{t+1,i,o} \widetilde{H}_{ti}, \quad (21)$$

$$\widetilde{H}_{ti} = \sum_{j:(j,i) \in \mathcal{U}_t^S} \widehat{X}_{tji} \widehat{X}_{toj} w_{tji} - \sum_{j:(i,j) \in \mathcal{U}_t^S} \widehat{X}_{toi} w_{tij}, \quad \forall i \in C, \quad (22)$$

where  $[w_{tij}]_{(i,j) \in \mathcal{U}_t^S} = f_S(\mathcal{G}_{1:t-1}^P, [\widehat{X}_{tij}]_{(i,j) \in \mathcal{U}_t^S}; \theta_k^S)$ , and  $\mathcal{B} \subset \mathcal{T}_k^{S,\text{tr}}$  is a batch. For each  $k \in [1 : n_{\text{fit}}]$ , the best hyperparameter set is selected from among  $n_{\text{hyper}}$  configurations based on performance on  $\mathcal{T}_k^{S,\text{va}}$ . Loss (19) maximizes a squared estimate of (12) when  $\hat{\mu}_G > 0$ , where  $\mathbb{I}$  denotes the indicator function. Since maximizing  $\hat{\mu}_G / \hat{\sigma}_G^2$  when  $\hat{\mu}_G \leq 0$  can increase the loss, we instead maximize the estimated profit. For  $f_S$  to learn from realized samples, during training and evaluation, we use  $\widetilde{X}_{t+1,i,o}, \widetilde{X}_{tij}, \widetilde{Y}_{t,1,i}, \widetilde{Y}_{t,1,o}$  in lieu of  $X_{t+1,i,o}, X_{tij}, Y_{t,1,i}^f, Y_{t,1,o}^f$  in (13) and (14), yielding (21) and (22), respectively. Note that  $\widetilde{Y}_{t,1,i}, \widetilde{Y}_{t,1,o}, \widetilde{X}_{t+1,i,o}$ , and  $\widetilde{X}_{tji}$  are not used by  $f_S$  during inference—unlike in prior studies [25, 26, 34].

To handle (15)–(18), we transform (14) and (15) into empirical constraints and define  $h_{SO}$ . Replacing  $\mathbb{E}X_{tij}$  with the predictions  $\widehat{X}_{tij}$  from  $f_P$ , we convert the expectation of (14) and (15) into

$$\widehat{H}_{ti} = \sum_{j:(j,i) \in \mathcal{U}_t^S} \widehat{X}_{tji} \widehat{X}_{toj} w_{tji} - \sum_{j:(i,j) \in \mathcal{U}_t^S} \widehat{X}_{toi} w_{tij}, \quad \forall i \in C, \quad (23)$$

$$\widehat{H}_{ti} = 0, \quad \forall i \in C \setminus \{o\}, \quad (24)$$

respectively. While (16), (18) and (24) can be enforced by projecting the outputs of  $g_S$  onto their linear constraint space, (17) is nonlinear. Thus, we introduce  $u_{tij}$  subject to

$$\sum_{j:(i,j) \in \mathcal{U}_t^S} u_{tij} = 0, \quad \forall i \in C \setminus \{o\}, \quad (25)$$

$$\widehat{X}_{toi} u_{tij} + \widehat{X}_{toj} \widehat{X}_{tji} u_{tji} = 0, \quad \forall (i,j) \in \mathcal{U}_t^S : i < j, \quad (26)$$

$$u_{tij} \in \mathbb{R}, \quad \forall (i,j) \in \mathcal{U}_t^S, \quad (27)$$

$$w_{tij} = \frac{[u_{tij}]^+}{\sum_{(i,j) \in \mathcal{U}_t^S} [u_{tij}]^+}, \quad \forall (i,j) \in \mathcal{U}_t^S, \quad (28)$$

if  $\sum_{(i,j) \in \mathcal{U}_t^S} [u_{tij}]^+ > 0$ . Then, due to (26) and (28),  $w_{tij}$  satisfies (17). In addition, (28) enforces (16) and (18), and (25) enforces (24). The details are provided in the proof of Proposition 1. We let  $\widehat{D}_t = \{[u_{tij}]_{(i,j) \in \mathcal{U}_t^S} : (25)–(27)\} \subset \mathbb{R}^{|\mathcal{U}_t^S|}$  and let  $\widehat{B}_t$  denote a matrix whose columns form a basis for  $\widehat{D}_t$ . We then define  $h_{SO}([u'_{tij}]_{(i,j) \in \mathcal{U}_t^S}) = [w_{tij}]_{(i,j) \in \mathcal{U}_t^S}$  as in (28), where

$$[u_{tij}]_{(i,j) \in \mathcal{U}_t^S} = \text{Proj}_{\widehat{D}_t} [u'_{tij}]_{(i,j) \in \mathcal{U}_t^S} = \widehat{P}_t [u'_{tij}]_{(i,j) \in \mathcal{U}_t^S}, \quad (29)$$

if  $\sum_{(i,j) \in \mathcal{U}_t^S} [u_{tij}]^+ > 0$ . Here,  $\widehat{P}_t = \widehat{B}_t (\widehat{B}_t^T \widehat{B}_t)^{-1} \widehat{B}_t^T$  is the projection matrix onto  $\widehat{D}_t$ . Since the projection in (29) implies  $[u_{tij}]_{(i,j) \in \mathcal{U}_t^S}$  satisfies (25)–(27), the proposition below shows  $h_{SO}$  guarantees the empirical constraints (24) and (16)–(18), which correspond to (15)–(18). Unlike [34], our method does not allow constraint violations.

**PROPOSITION 1.** *Equality  $C_1 = C_2$  holds if  $\widehat{X}_{tij} \widehat{X}_{tji} = 1$  for all  $(i,j) \in \mathcal{U}_t^S$ , where*

$$C_1 = \{[w_{tij}]_{(i,j) \in \mathcal{U}_t^S} : (16)–(18) \text{ and } (24)\},$$

$$C_2 = \{[w_{tij}]_{(i,j) \in \mathcal{U}_t^S} : (25)–(28), \sum_{(i,j) \in \mathcal{U}_t^S} [u_{tij}]^+ > 0\}.$$

**PROOF.** First, we show  $C_1 \subseteq C_2$ . Let  $[w_{tij}]_{(i,j) \in \mathcal{U}_t^S} \in C_1$ . We define

$$u_{tij} = w_{tij} \mathbb{I}_{\{w_{tij} > 0\}} - \frac{\widehat{X}_{toj} \widehat{X}_{tji}}{\widehat{X}_{toi}} w_{tji} \mathbb{I}_{\{w_{tji} \leq 0\}}. \quad (30)$$

Then, (27) holds. By (16) and (17), there exists  $(i,j)$  such that  $w_{tij} > 0$  and  $w_{tji} = 0$ , implying  $u_{tij} > 0$ . Thus,  $\sum_{(i,j) \in \mathcal{U}_t^S} [u_{tij}]^+ > 0$ . From (30) and (18),  $[u_{tij}]^+ = w_{tij}$ , which yields (28) by (16).

By (23),  $\widehat{H}_{ti} = \sum_{j:(i,j) \in \mathcal{U}_t^S} (\widehat{X}_{tji} \widehat{X}_{toj} w_{tji} - \widehat{X}_{toi} w_{tij})$  since  $(i,j) \in \mathcal{U}_t^S$  if and only if  $(j,i) \in \mathcal{U}_t^S$ . Since  $w_{tij} \geq 0$ , we have  $\widehat{H}_{ti} = \sum_{j:(i,j) \in \mathcal{U}_t^S} (\widehat{X}_{tji} \widehat{X}_{toj} w_{tji} \mathbb{I}_{\{w_{tji} > 0\}} - \widehat{X}_{toi} w_{tij} \mathbb{I}_{\{w_{tij} > 0\}})$ . Since  $w_{tji} \cdot \mathbb{I}_{\{w_{tji} > 0\}} = w_{tji} \mathbb{I}_{\{w_{tji} \leq 0\}}$ , by (17) and (18), we obtain  $\widehat{H}_{ti} = \sum_{j:(i,j) \in \mathcal{U}_t^S} (\widehat{X}_{tji} \widehat{X}_{toj} w_{tji} \mathbb{I}_{\{w_{tji} \leq 0\}} - \widehat{X}_{toi} w_{tij} \mathbb{I}_{\{w_{tij} > 0\}}) = -\widehat{X}_{toi} \cdot \sum_{j:(i,j) \in \mathcal{U}_t^S} u_{tij}$ , which, together with (24), yields (25).

By (30), since  $\widehat{X}_{tij} \widehat{X}_{tji} = 1$ , we obtain  $\widehat{X}_{toi} u_{tij} + \widehat{X}_{toj} \widehat{X}_{tji} u_{tji} = \widehat{X}_{toi} w_{tij} (\mathbb{I}_{\{w_{tij} > 0\}} - \mathbb{I}_{\{w_{tji} \leq 0\}}) + \widehat{X}_{toj} \widehat{X}_{tji} w_{tji} (\mathbb{I}_{\{w_{tji} > 0\}} - \mathbb{I}_{\{w_{tij} \leq 0\}})$ . By (17) and (18),  $w_{tij} (\mathbb{I}_{\{w_{tij} > 0\}} - \mathbb{I}_{\{w_{tji} \leq 0\}}) = w_{tji} (\mathbb{I}_{\{w_{tji} > 0\}} - \mathbb{I}_{\{w_{tij} \leq 0\}}) = 0$ . Thus, we have (26). Hence,  $C_1 \subseteq C_2$ .

Second, we show  $C_2 \subseteq C_1$ . Let  $[w_{tij}]_{(i,j) \in \mathcal{U}_t^S} \in C_2$ . From the definition of  $C_2$ , there exists  $[u_{tij}]_{(i,j) \in \mathcal{U}_t^S}$  such that  $u_{tij}$  defines

$w_{tij}$ , (25) to (28) hold, and  $\sum_{(i,j) \in \mathcal{U}'_t} [u_{tij}]^+ > 0$ . For  $i \in C \setminus \{o\}$ ,

$$\widehat{H}_{ti} \sum_{(i,j) \in \mathcal{U}'_t} [u_{tij}]^+ \quad (31)$$

$$= \sum_{j:(j,i) \in \mathcal{U}'_t} \widehat{X}_{tji} \widehat{X}_{toj} \left[ \frac{-\widehat{X}_{toi} u_{tij}}{\widehat{X}_{tji} \widehat{X}_{toj}} \right]^+ - \sum_{j:(i,j) \in \mathcal{U}'_t} \widehat{X}_{toi} [u_{tij}]^+ \quad (32)$$

$$= -\widehat{X}_{toi} \sum_{j:(i,j) \in \mathcal{U}'_t} ([u_{tij}]^+ - [-u_{tij}]^+) \quad (33)$$

$$= -\widehat{X}_{toi} \sum_{j:(i,j) \in \mathcal{U}'_t} u_{tij}. \quad (34)$$

We obtain (32) from (23), (26) and (28). The definition of  $[\cdot]^+$  yields (33) and (34). We have  $\widehat{H}_{ti} \sum_{(i,j) \in \mathcal{U}'_t} [u_{tij}]^+ = -\widehat{X}_{toi} \sum_{j:(i,j) \in \mathcal{U}'_t} u_{tij} = 0$ , by (25). Since  $\sum_{(i,j) \in \mathcal{U}'_t} [u_{tij}]^+ > 0$ , (24) holds.

By (28), (16) holds. By combining (26), i.e.,  $u_{tji} u_{tji} \leq 0$ , and (28), we obtain (17). From (28), (18) holds. Hence,  $C_2 \subseteq C_1$ .  $\square$

We define  $f_{SI}(\mathcal{G}_{1:t-1}^P, [\widehat{X}_{t,ij}]_{(i,j) \in \mathcal{U}'_t}) = \mathcal{G}_{t-1}^{P''} = (\mathcal{U}'_t, \widehat{A}_t, [\widehat{\alpha}_{tij}]_{(i,j) \in \mathcal{U}'_t}, [\widehat{\beta}_{tij}]_{(i,j) \in A_t})$ . Here,  $\widehat{\alpha}_{tij} = [\widehat{\alpha}_{tijt'}]_{t' \in \mathcal{T}^R} \in \mathbb{R}^{|\mathcal{T}^R|}$  where  $\widehat{\alpha}_{tijt'} = \frac{1}{|\mathcal{T}'(i,j,t,t')|} \sum_{t \in \mathcal{T}'(i,j,t,t')} \widehat{\alpha}_{tij}$ ,  $\widehat{\alpha}_{tij} = \log \widehat{X}_{tij} - \log \widehat{V}_{ti} + \log \widehat{V}_{tj}$ , and  $[\log \widehat{V}_{ti}]_{i \in C}$  is the least-squares solution to (8) and (9) when  $\widehat{V}_{ti}, \widehat{V}_{tj}, \widehat{X}_{tij}$  are replaced by  $\widehat{V}_{ti}, \widehat{V}_{tj}, \widehat{X}_{tij}$ , respectively. We define  $\mathcal{T}'(i,j,t,t') = \{t'' \in [t-t'+1:t] : (i,j) \in L_{t''}\}$  as the set of dates  $t'' \in [t-t'+1:t]$  on which both exchanges  $(i,j)$  and  $(j,i)$  are directly tradable. As discussed in (8) and (9),  $\widehat{\alpha}_{tij}$  provides information about estimated arbitrages, and  $\widehat{\alpha}_{tij}$  consists of temporal averages of  $\widehat{\alpha}_{tij}$  across different look-back windows in  $\mathcal{T}^R$ . We define  $\widehat{A}_t = \{(i,j) : |\widehat{\beta}_{tij}| > \epsilon_S\}$ , where  $\widehat{\beta}_{tij}$  is the  $(i,j)$ -th entry of  $\widehat{P}_t$ , and  $\epsilon_S > 0$  is a small number. We define  $\widehat{\beta}_{tij} = [\widehat{\beta}_{tijt'}]_{t' \in \mathcal{T}^R} \in \mathbb{R}^{|\mathcal{T}^R|}$  where  $\widehat{\beta}_{tijt'} = \frac{1}{|\mathcal{T}'(i,j,t,t')|} \sum_{t \in \mathcal{T}'(i,j,t,t')} \widehat{\beta}_{tij}$ . The edge features provide information about how much the target node's trading quantity changes as the source node's quantity changes. We define  $g_S(\mathcal{G}_{t-1}^{P''})$  with  $L$  layers as  $\text{SLP}_H(g_{CN,L}(g_{CS}(\mathcal{G}_{t-1}^{P''})))$ .

## 5 Experiments

We use Finaeon data [14] from Jan. 1, 1995 to Dec. 31, 2024, covering ten of the most traded currencies in 2022 [6]: USD, EUR, JPY, GBP, AUD, CAD, CHF, HKD, SGD, and SEK. CNY is excluded due to incomplete data. The data include daily closing exchange rates among the ten currencies and their IRs for 1-, 2-, 5-, and 10-year government bonds. To satisfy assumption (1), FX rates  $\widehat{X}_{tij}$ , excluding direct arbitrages, are redefined as the geometric average of  $\widehat{X}_{tij}$  and  $1/\widehat{X}_{tji}$ . For temporal stability, missing values in  $\widehat{X}_{tij}$  are forward-filled for up to seven days only when computing (8) and (9). Missing IR values are forward-filled for up to 30 days; any remaining missing values are imputed via regression on log(maturity).

To ensure data stability, we apply the following preprocessing. FX rates involving JPY and SEK that are mis-scaled by factors of 10, 100, or 10,000 are rescaled to their correct magnitudes. Erroneous values—those extremely large or small relative to adjacent time-series entries—are removed: SGDAUD has repeated entries from Dec. 29, 2014 to May 20, 2015; SEKEUR falls below 0.069 on Jun. 7, 2024; and AUDCHF exceeds 0.85 on Apr. 24, 2023, Jan. 12, 2024, and

Jan. 26, 2024. These values are identified by comparison with historical charts from Yahoo Finance [33]. IR entries above 90% for HKG and SGP during 2022 and 2023 are removed, following [28, 29].

Trading is assumed to occur only on weekdays, with weekends excluded from trading dates. To ensure at least a few thousand training graphs for GL models, we define  $t_1$  and  $t_k$  as the first trading days of 2015 and each subsequent quarter, respectively, and set  $n_{\text{fit}} = 40$  and  $n_{\text{sy}} = 5$ . We conduct three FXSA experiments by setting each of the most-traded currencies USD, EUR, and JPY as home currency  $o$ . We estimate  $\widehat{Y}_{t,1,i}$  by dividing the 1-year government bond rate of  $i$  by 365. We set  $\mathcal{T}^R = \{1, 3, 5, 10, 15, 20\}$ . We set  $n_{\text{hyper}} = 9$  and consider hyperparameter configurations  $(\#params, L) \in \{10^4, 5 \cdot 10^4, 10^5\} \times \{2, 3, 4\}$ , where  $\#params$  is the number of parameters in  $g_P$  and  $g_S$ , and  $L$  is the number of layers. Each SLP in  $g_P$  and  $g_S$  uses the same number of hidden neurons with LeakyReLU activation. Our implementation uses Python, Gurobi, R, along with PyTorch, PyTorch-Geometric, PuLP, SciPy, NumPy, Pandas, and lawstat libraries. Experiments are run on a workstation with i9-14900KF, 96GB RAM, and RTX 4070 Ti SUPER.

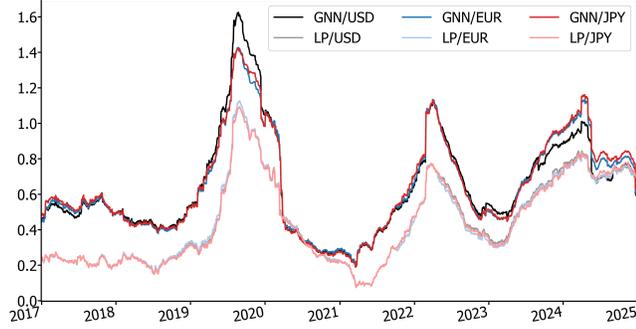
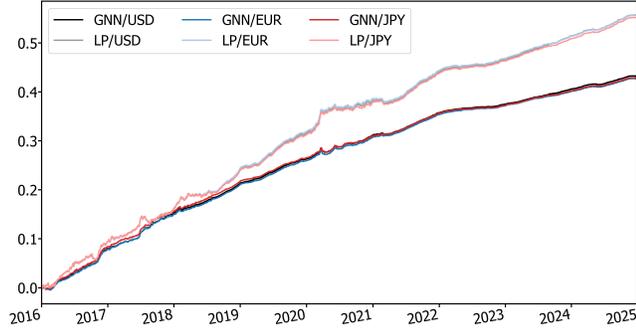
**Foreign Exchange Statistical Arbitrage.** In Table 1, our FXSA method (GNN) achieves, on average, 61.89% higher information ratio and 45.51% higher Sortino ratio than the benchmark LP (3)–(6), consistently outperforming across all three home currencies. The metrics where our method outperforms on average are in bold. Both our method and the LP use the same predictions for all  $\widehat{X}_{tij}$ , obtained from the GNN with FX/IR reported in Table 2, which we discuss shortly. Our method maintains a consistently higher information ratio than the LP throughout most of the test period (Figure 1), indicating superior risk-adjusted performance. Additionally, the sharp drops in 2020, 2022, and 2024 (Figure 1) likely account for geopolitical shocks unseen in the training data: the COVID-19 pandemic, the Ukraine–Russia war, and the war in the Middle East, respectively. Although the average annual return is slightly lower, GNN shows a steadier upward trend than LP (Figure 2).

The superior performance of our FXSA method stems from better risk management, in our opinion. By aiming to maximize the information ratio in (19), the method considers both return and risk. As a result, our approach achieves, on average, 52.23% lower annual volatility and 44.77% lower maximum drawdown (MDD) (Table 1). The approach also maintains 30.09% lower average holdings,  $\sum_{i \in C} |\widehat{H}_{ti} \widehat{X}_{tio}|$ . Our method consistently exhibits lower values than LP throughout most of the test period (Figure 3), thereby reducing exposure to the risk of FX rate movements. Although both methods enforce zero expected holdings (i.e., (4) for LP and (25)–(27) with Proposition 1 for our method), realized values may deviate; our method keeps them closer to zero by minimizing (19). Lastly, the proposed method shows an 83.12% lower average Herfindahl-Hirschman index (HHI), indicating more diversified trading and thus reduced risk. These outperformance results are consistent across the three home currencies.

**Foreign Exchange Rate Prediction.** Table 2 shows our FXRP method outperforms the benchmark MLP in mean squared error (MSE). FX, CV and IR correspond to  $\mathbf{x}_{tij}$ ,  $\mathbf{v}_{ti}$ , and  $\mathbf{y}_{ti}$  features in  $\mathcal{G}_t^{PF}$ , respectively. For each  $\mathcal{T}_k^{\text{te}}$  ( $k \in [1, n_{\text{fit}}]$ ), we compute an MSE, and each value in the second and third rows reports the average

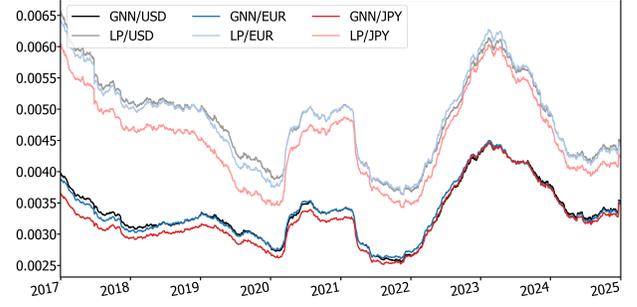
**Table 1: Evaluation Metrics of Trading Strategies (in %)**

Model	LP			GNN			Change Avg.
	USD	EUR	JPY	USD	EUR	JPY	
Info. Ratio	27.21	27.10	26.87	43.86	43.54	44.01	<b>61.89</b>
Sort. Ratio	32.70	32.46	32.53	47.44	47.67	47.03	<b>45.51</b>
Return	6.19	6.18	6.13	4.80	4.74	4.75	-22.73
Volatility	1.41	1.41	1.41	0.68	0.67	0.67	<b>-52.23</b>
MDD	1.61	1.60	1.72	0.92	0.89	0.91	<b>-44.77</b>
HHI	31.60	31.57	31.66	5.38	5.32	5.31	<b>-83.12</b>
Holding	0.50	0.49	0.46	0.34	0.34	0.33	<b>-30.09</b>

**Figure 1: 365-Day Rolling Information Ratio****Figure 2: Cumulative Sum of P&L ( $G_t$ ), in Units of  $o$** 

of the MSEs across all  $k$  for the corresponding model trained with the feature combination. All GNN results outperform their MLP counterparts, with the worst GNN outperforming the best MLP.

The outperformance of our GNN is statistically validated. In Table 2, each value in the fourth to ninth rows is the p-value of a statistical test comparing the GNN and MLP under the same feature set. Wilcoxon, KS, and symmetry refer to the Wilcoxon signed-rank test, Kolmogorov–Smirnov test, and the symmetry test from [19], respectively. The paired t-, permutation, and Wilcoxon tests evaluate the one-sided hypothesis  $H_0 : \text{MSE}(\text{MLP}) \leq \text{MSE}(\text{GNN})$  versus  $H_1 : \text{MSE}(\text{MLP}) > \text{MSE}(\text{GNN})$ , where the paired t-test and permutation test assess mean differences, and the Wilcoxon test assesses median differences. To validate the assumptions behind these tests, we use the Shapiro–Wilk, KS, and symmetry tests. Bold

**Figure 3: 365-Day Rolling Avg. of  $\sum_{i \in C} |\tilde{H}_{ti} \tilde{X}_{tio}|$ , in Units of  $o$** **Table 2: MSE Comparison Between GNN and MLP Across Feature Combinations**

		FX	FX/CV	FX/IR	FX/IR/CV
Avg. MSE (Unit: $10^{-5}$ )	GNN	2.8680	2.8648	<b>2.8635</b>	2.8650
	MLP	2.9112	2.8704	2.9083	2.8710
	Avg.	2.8896	2.8676	2.8859	2.8680
P-Value	Paired T	0.0000	0.0082	0.0000	<b>0.0054</b>
	Wilcoxon	<b>0.0000</b>	<b>0.0112</b>	<b>0.0000</b>	0.0045
	Permutation	0.0000	0.0067	0.0000	0.0052
	Shapiro	0.0025	0.0004	0.0013	0.2863
	KS	0.2390	0.1154	0.0709	0.7923
	Symmetry	0.1440	0.0620	0.0880	0.5600

p-values indicate statistically significant results (at the 0.05 level) where test assumptions hold. Thus, our GNN significantly outperforms the MLP across all feature sets. Moreover, Table 3 shows the GNNs statistically significantly outperform the MLPs across all feature combinations, except when using FX features alone.

Table 2 demonstrates the necessity of comprehensive relational information among FX rates and IRs captured by GNNs. Since the CV features are derived from multiple FX differences in (8) and (9), they provide some relational information about FX rates, which may explain why the MLP with FX/IR/CV outperforms the MLP with FX/IR. However, CV captures limited relational information compared to the GNNs, as GNNs across all feature combinations outperform the MLP with FX/IR/CV. Moreover, the GNN with FX/IR/CV underperforms the GNN with FX/IR, indicating that CV does not provide additional useful information beyond what the GNN with FX/IR captures. Lastly, the best performance of the GNN with FX/IR underscores the importance of its information about complex multi-currency-IR relationships.

## 6 Conclusion

This paper proposes a two-stage GL approach to generate FXSAs, addressing two key limitations in the literature: (i) the absence of GL methods for FXRP that leverage multi-currency and currency-IR relationships, and (ii) the disregard of the time lag between price observation and trade execution. In the first stage, we formulate FXRP as an edge-level regression problem on a discrete-time spatiotemporal graph to capture complex multi-currency and

**Table 3: P-Values for MSE Comparison Between GNN and MLP Across Feature Combinations, Testing  $H_0 : \text{MSE}(MLP) \leq \text{MSE}(GNN)$  vs.  $H_1 : \text{MSE}(MLP) > \text{MSE}(GNN)$ . Final letters T, W, and P indicate the paired t-, Wilcoxon signed-rank, and permutation tests, respectively, applied when their assumptions hold.**

		GNN			
		FX	FX/CV	FX/IR	FX/IR/CV
MLP	FX	0.000 W	0.000 W	0.000 W	0.000 W
	FX/CV	0.147 P	0.011 W	0.001 W	0.003 W
	FX/IR	0.000 W	0.000 W	0.000 W	0.000 P
	FX/IR/CV	0.113 W	0.005 P	0.000 W	0.005 T

currency-IR relationships. We present a GL method to address the FXRP problem, leveraging the spatiotemporal graph defined with currencies and exchanges as nodes and edges, respectively, where IRs serve as node features, and FX rates serve as edge features. In the second stage, we present a stochastic optimization problem to exploit FXSAs while accounting for the observation-execution time lag. To solve this problem, we propose a GL method that enforces constraints through projection and ReLU to maximize the information ratio, while utilizing the predictions from the FXRP method for the constraint parameters and node features. We prove our GL method satisfies the empirical arbitrage constraints. Our GL method for FXSA uses a graph in which nodes represent exchanges and edges encode their influences. Experimental results show our FXRP method achieves statistically significant improvements over non-graph methods. Moreover, our FXSA method achieves a 61.89% higher information ratio and a 45.51% higher Sortino ratio than the benchmark. Our approach presents a novel perspective on FX prediction and StatArbs through the lens of GL.

## References

- [1] S. Ahmed, S.-U. Hassan, N. R. Aljohani, and R. Nawaz. 2020. FLP-LSTM: A novel prediction system using FOREX loss function. *Applied Soft Computing* 97 (2020), 106780.
- [2] Y. Aiba, N. Hatano, H. Takayasu, K. Marumo, and T. Shimizu. 2002. Triangular arbitrage as an interaction among foreign exchange rates. *Physica A: Statistical Mechanics and its Applications* 310, 3 (2002), 467–479.
- [3] O. Ajumi and A. Kaushik. 2017. Exchange rates prediction via deep learning and machine learning: A literature survey on currency forecasting. *International Journal of Science and Research* 7 (2017).
- [4] C. Amat, T. Michalski, and G. Stoltz. 2018. Fundamentals and exchange rate forecastability with simple machine learning methods. *Journal of International Money and Finance* 88 (2018), 1–24.
- [5] M. Ayitey Junior, P. Appiahene, O. Appiah, and C. N. Bombie. 2023. FOREX market forecasting using machine learning: Systematic literature review and meta-analysis. *Journal of Big Data* 10, 1 (2023), 9.
- [6] Bank for International Settlements. 2022. OTC foreign exchange derivatives by instrument, counterparty and maturity. [https://data.bis.org/topics/DER/tables-and-dashboards/BIS\\_DER\\_D11\\_3\\_1.0](https://data.bis.org/topics/DER/tables-and-dashboards/BIS_DER_D11_3_1.0). Accessed: Jul. 13, 2025.
- [7] R. B.apat. 2014. *Graphs and Matrices*. Springer London, Chapter 2, 14. Lemma 2.2.
- [8] W. Cao, W. Zhu, W. Wang, Y. Demazeau, and C. Zhang. 2020. A deep coupled LSTM approach for USD/CNY exchange rate forecasting. *IEEE Intelligent Systems* 35, 2 (2020), 43–53.
- [9] A. Carriero, G. Kapetanios, and M. Marcellino. 2009. Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting* 25, 2 (2009), 400–417.
- [10] Z. Cui, W. Qian, S. Taylor, and L. Zhu. 2020. Detecting and identifying arbitrage in the spot foreign exchange market. *Quantitative Finance* 20, 1 (2020), 119–132.
- [11] B. J. de Almeida, R. F. Neves, and N. Horta. 2018. Combining support vector machine with genetic algorithms to optimize investments in FOREX markets with high leverage. *Applied Soft Computing* 64 (2018), 596–613.
- [12] W. Du, A. Tepper, and A. Verdelhan. 2018. Deviations from covered interest rate parity. *The Journal of Finance* 73, 3 (2018), 915–957.
- [13] D. J. Fenn, S. D. Howison, M. McDonald, S. Williams, and N. F. Johnson. 2009. The mirage of triangular arbitrage in the spot foreign exchange market. *International Journal of Theoretical and Applied Finance* 12, 08 (2009), 1105–1123.
- [14] Finaeon. 2025. Global financial database (GFDatabase). <https://www.finaeon.com>. Accessed: Apr. 17, 2025.
- [15] S. Galeshchuk and S. Mukherjee. 2017. Deep networks for predicting direction of change in foreign exchange rates. *Intelligent Systems in Accounting, Finance and Management* 24, 4 (2017), 100–110.
- [16] M. S. Islam and E. Hossain. 2021. Foreign exchange currency rate prediction using a GRU-LSTM hybrid network. *Soft Computing Letters* 3 (2021), 100009.
- [17] A. Langenohl. 2018. Sources of financial synchronism: Arbitrage theory and the promise of risk-free profit. *Finance and Society* 4, 1 (2018), 26–40.
- [18] S. Mahmoodzadeh and M. C. Tseng. 2020. Spot arbitrage in FX market and algorithmic trading: Speed is not of the essence. *Market Microstructure and Liquidity* 6, 01n04 (2020), 2050011.
- [19] W. Miao, Y. R. Gel, and J. L. Gastwirth. 2006. A new test of symmetry about an unknown median. In *Random walk, sequential analysis and related topics: A festschrift in honor of Yuan-Shih Chow*. World Scientific, 199–214.
- [20] M. Pahlavani and R. Roshan. 2015. The comparison among ARIMA and hybrid ARIMA-GARCH models in forecasting the exchange rate of Iran. *International Journal of Business and Development Studies* 7, 1 (2015), 31–50.
- [21] J. F. Pfahler. 2022. Exchange rate forecasting with advanced machine learning methods. *Journal of Risk and Financial Management* 15, 1 (2022), 2.
- [22] D. Pradeepkumar and V. Ravi. 2014. FOREX rate prediction using chaos, neural network and particle swarm optimization. In *International Conference in Swarm Intelligence*. Springer, 363–375.
- [23] M. Rupasinghe, M. N. Halgamuge, and N. T. Q. Vinh. 2019. Forecasting trading-time based profit-making strategies in FOREX industry: Using Australian FOREX data. In *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 1–9.
- [24] W. F. Sharpe. 1998. The Sharpe ratio. *Streetwise—the Best of the Journal of Portfolio Management* 3, 3 (1998), 169–85.
- [25] W. Soon and H.-Q. Ye. 2007. Currency arbitrage detection using a binary integer programming model. In *2007 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, Piscataway, NJ, USA, 867–870.
- [26] K. Tatsumura, R. Hidaka, M. Yamasaki, Y. Sakai, and H. Goto. 2020. A currency arbitrage machine based on the simulated bifurcation algorithm for ultrafast detection of optimal opportunity. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, Piscataway, NJ, USA, 1–5.
- [27] M. P. Taylor. 1995. The economics of exchange rates. *Journal of Economic Literature* 33, 1 (1995), 13–47.
- [28] The Association of Banks in Singapore. 2025. Singapore benchmark rates (SORA, SIBOR, SOR). <https://abs.org.sg/benchmark-rates>. Accessed: Jul. 16, 2025.
- [29] The Hong Kong Association of Banks. 2025. Hong Kong interbank offered rate (HIBOR). <https://www.hkab.org.hk/en/rates/hibor>. Accessed: Jul. 16, 2025.
- [30] B. Tuckman and P. Porfirio. 2003. Interest rate parity, money market basis swaps, and cross-currency basis swaps. *Fixed Income Liquid Markets Research, Lehman Brothers* 1 (2003).
- [31] U. Ullah and Z. Huang. 2025. Forecasting currency exchange rate through a hybrid time series analysis. *IEEE Transactions on Consumer Electronics* (2025), 1–1.
- [32] H. R. Varian. 1987. The arbitrage principle in financial economics. *Journal of Economic Perspectives* 1, 2 (1987), 55–72.
- [33] Yahoo Finance. 2025. Historical exchange rate charts. <https://finance.yahoo.com>. Accessed: Jul. 16, 2025.
- [34] D. Zhang. 2025. Efficient triangular arbitrage detection via graph neural networks. arXiv:2502.03194
- [35] Y. Zhang and S. Hamori. 2020. The predictability of the exchange rate when combining machine learning and fundamental models. *Journal of Risk and Financial Management* 13, 3 (2020), 48.
- [36] Y. Zhao and M. Khushi. 2020. Wavelet Denoised-ResNet CNN and LightGBM method to predict FOREX rate of change. In *2020 International Conference on Data Mining Workshops (ICDMW)*. IEEE, 385–391.