FX

A dataset contains day-level price of 21 currencies in exchange of US dollar with a total of 65478 data samples.

Dataset Snapshot

NATURE OF CONTENT
Date, trading price (open, high, low, close price), currency ticker and day of the week (0, 1, 2, 3, 4 representing Monday to Friday).

BREAKDOWN-BY INSTANCE

Total instances	65478
Training	52416
Validation	6531
Testing	6531
Total cryptocurrencies	21
Instances per stock	3118

NOTES
Stock data is collected daily from 2009-01-02 to 2020-01-01 on all trading days. The recommend split is [0.8,0.1,0.1] for training, validation and testing respectively.

EXAMPLES OF ACTUAL DATA POINT

	date	open	high	low	close	tic	day
0	2009-01-02	0.705020	0.711997	0.692521	0.709320	AUDUSD=X	4
1	2009-01-02	0.430293	0.433050	0.427936	0.432339	BRLUSD=X	4
2	2009-01-02	0.818063	0.829050	0.814730	0.826856	CADUSD=X	4
3	2009-01-02	0.146994	0.146994	0.146994	0.146994	CNYUSD=X	4
4	2009-01-02	0.188094	0.188459	0.185815	0.187105	DKKUSD=X	4

Publisher & Licenses & Access

PUBLISHER
Federal Reserve's Data Program

DIRECT LINKS TO DATASET
link

LICENSE PERMISSIONS
You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

ACCESS
Open Access

ACCESS COST
Free

LICENSE TYPE(S)
CC0 1.0

Motivations & Use

DATASET PURPOSE
The dataset was created to provide representative data of currency trading for research in various quantitative trading tasks by selecting the mainstream countries' currencies.

INTENDED USE CASES
- Algorithmic trading
- Portfolio Management

EXTENDED USE
- Intraday trading
- High frequency trading

Collection

DATA SOURCE
Retrieved from Kaggle

DATA COLLECTION
We download the data from Kaggle following this link:
https://www.kaggle.com/datasets/dhruvildave/currency-exchange-rates

Preprocessing

INDICATOR ADJUSTMENT
The raw data consists of 8 indicators, which are date, open, high, low, close, adjcp, volume and tic. Our dataset uses adjusted close price (adjcp) to replace original close price because it is considered as a more accurate measure of cryptocurrency's value. Since the volume is not accessible, this indicator is also removed.

DATA CLEANING
Firstly, all the NaN terms are dropped, and it is observed that some of the cryptocurrencies are lack of data. Also, it can be observed that only very few days of data is missing. In order to maintain consistency, data of these dates are filtered out.

FEATURE GENERATION
We generate 11 temporal features to describe the financial markets. z_{open}, z_{high}, z_{low} represent the relative values of the open, high, low prices compared with the close price at current time step, respectively. z_{close} represents the relative values of the closing prices compared with time step t-1. z_{dk} represents a long-term moving average of the adjusted close prices during the last k time steps compared to the current close prices. The detailed calculation formulas are as follow:

Features	Calculation Formula
$z_{open}, z_{high}, z_{low}$ z_{close}, z_{adj_close}	$z_{open} = open_t / close_t - 1$ $z_{close} = close_t / close_{t-1} - 1$
$z_{d\_5}, z_{d\_10}, z_{d\_15}$ $z_{d\_20}, z_{d\_25}, z_{d\_30}$	$z_{d\_5} = \dfrac{\sum_{i=0}^{4} adj\_close_{t-i}/5}{adj\_close_t} - 1$

Maintenance & Status

STATUS	FIRST RELEASE	CURRENT VERSION
Actively Maintained	08/2022	1.0