



Market Selection Framework

Version 1.0.2

Last Revised: April 25, 2021

Contents

1	Introduction	2
2	Feature Descriptions	2
2.1	Technology	2
2.2	Legal and Compliance	3
2.3	Business Model	4
2.4	Data Availability	4
2.5	Price	5
2.6	Volume	6
2.7	Order Book	7
3	Feature Normalization	7
4	Rating Algorithm	8
5	Selection Algorithm	9

1 Introduction

The Coin Metrics Oversight Committee (the “Oversight Committee”) evaluates markets traded on digital asset exchanges as potential input data sources for the CM Reference Rates (the “Reference Rates”) using a Market Selection Framework. The framework consists of a fully-systematized process for evaluating markets. In this framework, a market refers to a specific traded asset pair on a specific exchange. Only spot markets are considered. It produces a unique set of candidate selected markets for each asset in the coverage universe.

The Oversight Committee evaluates new markets for inclusion as a selected market and assesses already selected markets using the Market Selection Framework on a quarterly basis and during interim periods if market conditions warrant. Such market conditions include, but are not limited to, material changes in an exchange’s solvency risk, material changes in the degree of free capital flows in and out of an exchange, the presence of long-lasting price differences from other exchanges, and during times of market stress.

2 Feature Descriptions

The Market Selection Framework consists of 45 features, 36 of which are currently implemented. Features represent individual measurable properties that provide an indication of the suitability for a market to serve as an input data source, which are combined to form a market rating.

Some of the features described in this document are indicator variables that encode qualitative information about a market or exchange. These indicator variables can require a degree of subjectivity in determining whether a market or exchange meets a certain criteria. In such cases, the indicator variable is set to true only if an unambiguous source is found that provides sufficient information to make an evaluation. In the absence of such a source, the indicator variable is set to false.

2.1 Technology

An assessment of whether the technology infrastructure of the market’s exchange provides sufficient availability and reliability for input data collection. Evaluates whether the exchange offers a REST API or websocket feed for data collection. Evaluates the performance of the API in terms of reliability.

1. **has_rest_api**: An indicator variable for the existence of a REST API.
2. **has_websocket_feed**: An indicator variable for the existence of a websocket feed.

3. **has_fix_gateway**: An indicator variable for the existence of a FIX gateway.
4. **has_historical_trade_data**: An indicator variable for whether the exchange offers historical trade data via its API.
5. **has_real_time_trade_data**: An indicator variable for whether the exchange offers real-time trade data via its API.
6. **has_real_time_order_book_data**: An indicator variable for whether the exchange offers real-time order book data via its API.
7. **api_downtime_incidents**: A feature that represents the stability of an API. Not yet implemented.

2.2 Legal and Compliance

An assessment of whether the market's exchange complies with applicable laws and regulations. Evaluates the exchange's legal risk exposure and whether it adheres to regulatory best practices. Evaluates whether the exchange has publicly-disclosed trading policies, uses market surveillance technology, and complies with national regulatory organizations, and enforces KYC and AML requirements. Evaluates whether the exchange has functioning fiat and digital asset withdrawals processed within a normal timeframe.

1. **has_trading_policy**: An indicator variable for whether the exchange has a trading policy to promote fair and transparent markets. The trading policy should explicitly address manipulative trading policies like front-running, wash trading, spoofing, layering, churning, and quote stuffing.
2. **has_market_surveillance**: An indicator variable for whether the exchange uses market surveillance technology to detect market manipulation practices, including front-running, wash trading, spoofing, layering, churning, and quote stuffing.
3. **has_regulatory_oversight**: An indicator variable for whether the exchange complies with applicable laws and regulations, including registering or getting licenses from national or regional regulatory organizations.
4. **has_fiat_insurance**: An indicator variable for whether the exchange maintains commercial insurance or is covered under government-provided insurance to insure against losses of customer funds denominated in fiat currencies.
5. **has_crypto_insurance**: An indicator variable for whether the exchange maintains commercial insurance to insure against losses of customer funds denominated in digital assets.

6. **has_kycaml**: An indicator variable for whether the exchange requires identity verification before being able to trade and follows know-your-customer (KYC) regulations and anti-money-laundering (AML) regulations. For the purposes of this indicator variable, an exchange which requires identity verification only if a customer wishes to deposit or withdraw fiat or if a customer wishes to withdraw digital assets is determined to not have sufficient controls. An exchange must comply with both KYC and AML regulations for this indicator variable to be true.
7. **has_free_capital_flows**: An indicator variable for whether the exchange has had a history of free capital flows over the last quarter, including functioning fiat and digital asset deposits and withdrawals that are processed within a normal timeframe.

2.3 Business Model

An assessment of the market's exchange with respect to its business model, including its fee structure and asset listing standards.

1. **has_fiat_markets**: An indicator variable that indicates whether the exchange has markets where the quote currency is a fiat currency.
2. **has_fees**: An indicator variable that indicates whether the exchange charges trading fees as a percentage of the trade size. Exchanges that charge zero fees or charge fees indirectly through a transaction mining model are determined to not charge fees.
3. **has_listing_standards**: An indicator variable that indicates whether the exchange has publicly disclosed a framework for deciding which assets to list.
4. **has_usa_hq**: An indicator variable for whether the company's headquarters are domiciled in the United States.

2.4 Data Availability

An assessment of the available data the market's exchange offers for the given digital asset, including the amount of historical data available for a market and the quoted currency of the market.

1. **market_days_history**: The number of days of historical data for the market.
2. **market_quote_modifier**: An optional modifier to give greater weight to a certain quote currency. Currently all quote currencies have equal weight.

2.5 Price

An assessment of the quality of the market's price data, including testing for the occurrence of price outliers and impactful price deviations from other markets, and implementing tests that determine whether the exchange's markets function as active markets in the underlying digital asset and are anchored by observable transactions entered into at arm's length between buyers and sellers.

In this section, prices are compared to the global median price. The global median price is defined as the median price of all markets in which the digital asset is the base currency from the following list of exchanges: ["Coinbase", "Poloniex", "Bittrex", "Gemini", "Kraken", "Binance", "Bitstamp", "itBit"]. This list of exchanges was selected by first calculating the the market ratings for each market but without the price-related features below. The median market rating was then calculated for each exchange and the top 8 exchanges were selected.

1. `market_open_mape_all`: The mean absolute percentage error of the market's daily open price compared to the global median's daily open price over the last 90 days.
2. `market_close_mape_all`: The mean absolute percentage error of the market's daily close price compared to the global median's daily close price over the last 90 days.
3. `market_low_mape_all`: The mean absolute percentage error of the market's daily low price compared to the global median's daily low price over the last 90 days.
4. `market_high_mape_all`: The mean absolute percentage error of the market's daily high price compared to the global median's daily high price over the last 90 days.
5. `market_open_mape_trimmed`: The mean absolute percentage error of the market's daily open price trimmed to exclude the bottom and top 5th percentiles compared to the global median's daily open price over the last 90 days.
6. `market_close_mape_trimmed`: The mean absolute percentage error of the market's daily close price trimmed to exclude the bottom and top 5th percentiles compared to the global median's daily close price over the last 90 days.
7. `market_low_mape_trimmed`: The mean absolute percentage error of the market's daily low price trimmed to exclude the bottom and top 5th percentiles compared to the global median's daily low price over the last 90 days.

8. `market_high_mape_trimmed`: The mean absolute percentage error of the market's daily high price trimmed to exclude the bottom and top 5th percentiles compared to the global median's daily high price over the last 90 days.

2.6 Volume

An assessment of the quality of the market's volume data, including testing for manipulated volume figures, and implementing tests that determine whether the exchange's markets function as active markets in the underlying digital asset and are anchored by observable transactions entered into at arm's length between buyers and sellers. The size of the exchange's markets is also considered.

1. `market_volume_usd`: The total volume of the market over the past 90 days in U.S. dollars.
2. `market_volume_dispersion`: The coefficient of variation of the market's daily volume in U.S. dollars over the past 90 days.
3. `market_volume_price_corr_raw`: The correlation of the market's daily return to detrended daily volume where volume is quoted in raw units over the past 90 days.
4. `market_volume_price_corr_usd`: The correlation of the market's daily return to detrended daily volume where volume is quoted in U.S. dollars over the past 90 days.
5. `alexa_rank`: The global rank of the exchange's website as reported by Alexa.
6. `alexa_page_views_per_million`: The average page views per million visitors of the exchange's website over the past month as reported by Alexa.
7. `alexa_reach_per_million`: The average reach per million visitors of the exchange's website over the past month as reported by Alexa.
8. `alexa_pvpmvu`: The total U.S. dollar volume of the exchange over the past month divided by the page views per million visitors as reported by Alexa.
9. `alexa_rpmvu`: The total U.S. dollar volume of the exchange over the past month divided by the reach per million visitors as reported by Alexa.
10. `similarweb_global_rank`: The global rank of the exchange's website as reported by SimilarWeb.
11. `similarweb_visit_monthly`: The number of monthly visits as reported by SimilarWeb.
12. `similarweb_vmvu`: The total U.S. dollar volume of the exchange over the past month divided by the monthly visits as reported by SimilarWeb.

2.7 Order Book

An assessment of the quality of the market's order book data, including tests for manipulated orders, and implementing tests that determine whether the market function as active markets in the underlying digital asset and are anchored by observable transactions entered into at arm's length between buyers and sellers. The liquidity of the market is also considered.

1. **order_book_depth**: The total volume of bids and offers on the order book within 1 percent of the mid price of the exchange's largest traded market where the given asset is the base currency reported in U.S. dollars. Not yet implemented.
2. **slippage**: The amount of slippage in percent terms if an immediate market sell order of \$50,000 U.S. dollars is executed of the exchange's largest traded market where the given asset is the base currency. Not yet implemented.
3. **spread**: The median of the spread calculated as the difference between the best bid minus the best ask divided by the mid-price over the past 30 days. Not yet implemented.
4. **order_book_depth_residual**: A regression model is fit by regressing volume on order book depth for the largest traded market for a collection of exchanges. Given an order book depth, an estimated volume is calculated and a residual is calculated as `actual_volume - estimated_volume`. Not yet implemented.
5. **slippage_residual**: A regression model is fit by regressing volume on slippage for the largest traded market for a collection of exchanges. Given a slippage, an estimated volume is calculated and a residual is calculated as `actual_volume - estimated_volume`. Not yet implemented.

3 Feature Normalization

All features are normalized to between 0 and 1, with a number closer to 1 meaning that the feature contributes positively to the likelihood that the market will be selected. The qualitative features are all indicator variables that take values 0 or 1. For quantitative features, a separate empirical cumulative distribution function is calculated for all the markets for each asset. The quantitative feature is normalized by converting the value to its equivalent value on the empirical cumulative distribution function.

4 Rating Algorithm

In order for a market to be eligible to receive a rating, the following three indicator variables for the market must be set to true: `has_rest_api`, `has_real_time_trade_data`, `has_real_time_order_book_data`. These indicator variables are required to be true because the ability to collect real-time trade data and real-time order book data via an API is necessary in order for the market to serve as an input data source.

The rating algorithm uses a weighted sum using a custom weighting function of the normalized features:

Feature	Weight
<code>has_websocket_feed</code>	1
<code>has_fix_gateway</code>	1
<code>has_historical_trade_data</code>	1
<code>has_trading_policy</code>	1
<code>has_market_surveillance</code>	1
<code>has_regulatory_oversight</code>	1
<code>has_kycaml</code>	1
<code>has_fiat_insurance</code>	1
<code>has_crypto_insurance</code>	1
<code>has_free_capital_flows</code>	3
<code>has_fiat_market</code>	1
<code>has_fees</code>	1
<code>has_listing_standards</code>	1
<code>has_usa_hq</code>	3
<code>market_days_history</code>	1
<code>market_quote_modifier</code>	1
<code>market_open_mape_all</code>	1
<code>market_close_mape_all</code>	1
<code>market_low_mape_all</code>	1
<code>market_high_mape_all</code>	1
<code>market_open_mape_trimmed</code>	1
<code>market_close_mape_trimmed</code>	1
<code>market_low_mape_trimmed</code>	1
<code>market_high_mape_trimmed</code>	1
<code>market_volume_usd</code>	3
<code>market_volume_dispersion</code>	1
<code>market_volume_price_corr_raw</code>	1
<code>market_volume_price_corr_usd</code>	1
<code>alexa_rank</code>	2
<code>alexa_page_views_per_million</code>	1
<code>alexa_reach_per_million</code>	1
<code>alexa_pvpmvu</code>	1
<code>alexa_rpmvu</code>	1

Feature	Weight
similarweb_global_rank	2
similarweb_visit_monthly	1
similarweb_vmvu	1

5 Selection Algorithm

Markets with a rating higher than 22.5 are selected. For each asset in the coverage universe, markets that are in the top 4 by rank are also selected, regardless of the market's rating. Any market with volume, measured in U.S. dollars over the past 90 days, of less than 1 percent of the volume of the selected market with the largest volume is excluded.