DARA Historical risk model description

Timestamping convention and data availability

Timestamp reflects the point in time such that all the data used by the model had been available before it. As calculations take time, model data delivery happens after that timepoint.

- `cutoff_time` field describes the end of an hourly semi-interval.
  For instance, `cutoff_time = 2024-02-03 05:05:00` corresponds to a semi-interval `[2024-02-03 04:05:00; 2024-02-03 05:05:00)`
  where the left band is included and the right is not.
- Intervals start at 5 min past an hour to avoid ambivalence of price action associated with funding fee payment.
- Timestamps are UTC (this matches exchange timestamps on e.g. Binance and OKX).
- Data is typically available 30 minutes after the corresponding interval ends, so e.g. for the `cutoff_time = 2024-02-03 05:05:00` interval
  the data would have been delivered around `2024-02-03 05:35:00`.

Factors description

Factor	Description	Factor values direction
momentum	Price momentum over medium term	Winners-minus-Losers past winners have high/positive momentum score, past losers - low/negative score
reversion	Price reversion over short term	Losers-minus-Winners past losers have high/positive reversion score, past winners - low/negative
beta	Sensitivity to market	High-minus-Low high beta tokens have high/positive beta score, low beta tokens - low/negative
size	SML based on liquidity proxy of size	Small-minus-Large small size tokens have high/positive size score, large size tokens - low/negative
trader_activity	Relative trading activity	Low-minus-High low trading activity tokens have high/positive activity score, high activity - low/negative
downside_risk	Downside volatility	High-minus-Low high downside volatility tokens have high/positive downside risk score, low volatility - low/negative
holder_sentiment	Time to close outstanding open interest at average trading rate	High-minus-Low tokens with long time to close open interest have high holder sentiment score, others - low/negative
volatility	Symmetric volatility	Low-minus-High low volatility tokens have high/positive volatility score, high volatility tokens - low/negative
market	Broad market	

Risk model files description

Contents	File name	Timestamping details
Factor loadings	exposures.parquet	`cutoff_time` timestamp indicates the end point for the latest hourly interval used to calculate loadings. E.g. `cutoff_time = 2024-02-03 05:05:00` corresponds to a semi-interval `[2024-02-03 04:05:00; 2024-02-03 05:05:00)`

Factor returns	factor_returns.parquet	Factor returns are calculated a posteriori, i.e. there is a 2 hour offset between factor returns and exposures data in case one needs to map it (e.g. to obtain idio returns).
		E.g. factor returns marked with
		`cutoff_time = 2024-02-03 05:05:00`
		correspond to factor exposures of semi-interval
		`[2024-02-03 02:05:00; 2024-02-03 03:05:00)`

Covariance estimation	estimated_covariance.parquet	Covariance is estimated based on the point-in-time returns (latest returns available at the corresponding timestamp) and is lookahead bias free.
		E.g. covariance marked with
		`cutoff_time = 2024-02-03 05:05:00`
		is based on historical returns up to the latest available semi-interval at the time
		`[2024-02-03 02:05:00; 2024-02-03 03:05:00)`

Important notes:

- Factor returns are given on a funding fee adjusted basis (include funding fee where applicable) and are cross-sectionally additive (i.e. if one wants to make returns serially additive they will need to convert given returns to log returns first).
- Factor return figures are provided for reference. Independent factor returns estimation is strongly advised. This will ensure that outlier handling policies and the choice of price benchmarks are both consistent with the rest of one's modelling process and provide better latency and granularity when needed.
- Covariance is given on hourly basis, i.e. one needs to adjust variances by 24 * 365 if they want annualized variance figures.