

- Risk. *See also* Hierarchical Risk Parity (HRP) approach; Strategy risk
 backtest statistics for uncovering, 195
 entropy application to portfolio concentration and, 276
 liquidity and, 7, 286
 ML algorithms for monitoring, 14
 PCA weights and, 35–36, 35*f*
 portfolio oversight and, 8
 profit taking and stop-loss limits and, 44
 structural breaks and, 249
 walk-forward (WF) approach and, 161
- Risk-based asset allocation approaches, 222
 HRP approach comparisons in, 236–238
 structural breaks and, 249
- Risk parity, 222. *See also* Hierarchical Risk Parity (HRP) approach
 HRP approach compared with traditional approach to, 231–232, 233*t*, 234, 235*f*
- Rolled prices, 37
- Roll model, 282–283
- Sample weights, 59–72
 average uniqueness of labels over lifespan and, 61–62, 61*f*
 bagging classifiers and uniqueness and, 62–63
 indicator matrix for, 64–65
 mean decrease accuracy (MDA) feature importance with, 116
 number of concurrent labels and, 60–61
 overlapping outcomes and, 59–60
 return attribution method and, 68–69
 sequential bootstrap and, 63–68
 time-decay factors and, 70–71, 72*f*
- Sampling features, 38–40
 downsampling and, 38
 event-based sampling and, 38–40, 40*f*
- Scalability
 bagging for, 101
 sample size in ML algorithms and, 38, 101
- Scikit-learn (sklearn)
 class weights in, 71
 cross-validation (CV) bugs in, 109–110
 grid search cross-validation in, 129–130
 labels and bug in, 55, 72, 94
 mean decrease impurity (MDI) and, 115
 neg log loss as scoring statistic and bug in, 134
 observation redundancy and bagging classifiers in, 97
 random forest (RF) overfitting and, 98–99
 support vector machine (SVM) implementation in, 101
 synthetic dataset generation in, 122
 walk-forward timefolds method in, 155
- Selection bias, 153–154, 167
- Sequential bootstraps, 63–68
 description of, 63–64
 implementation of, 64–65
 leakage reduction using, 105
 Monte Carlo experiments evaluating, 66–68, 68*f*
 numerical example of, 65–66
- Shannon, Claude, 263–264
- Sharpe ratio (SR) in efficiency measurements
 annualized, 205
 definition of, 203
 deflated (DSR), 204, 205*f*
 information ratio and, 205
 probabilistic (PSR), 203–204, 204*f*, 205–206, 218
 purpose of, 203
 targeting, for various betting frequencies, 212–213
- Shorting, in quantitative investing, 152

- Signal order flows, 295
- Simulations, overfitting of, 154
- Single feature importance (SFI), 117–118, 125–127, 126*f*
- Single future roll, 36–37
- Sklearn. *See* Scikit-learn
- Stacked feature importance, 121–122
- Standard bars (table rows), 26–28
 - dollar bars, 27–28, 28*f*, 44
 - purpose of, 26
 - tick bars, 26–27
 - time bars, 26, 43–44
 - volume bars, 27, 44
- Stationarity
 - data transformation method to ensure, 77–78
 - fractional differentiation applied to, 76–77
 - fractional differentiation
 - implementation methods for, 80–84
 - integer transformation for, 76
 - maximum memory preservation for, 84–85, 84*f*, 86*t*–87*t*
 - memory loss dilemma and, 75–76
- Stop-loss, and investment strategy exit, 211
- Stop-loss limits
 - asymmetric payoff dilemma and, 178–180
 - cases with negative long-run equilibrium and, 182–187, 186*f*, 187*f*–191*f*
 - cases with positive long-run equilibrium and, 180–182, 181*f*, 182*f*, 183*f*–186*f*
 - cases with zero long-run equilibrium and, 177–180, 177*f*, 178*f*, 179*f*
 - daily volatility computation and, 44–45
 - fixed-time horizon labeling method and, 44
 - investment strategies using, 170–171, 172, 211
 - learning side and size and, 48
 - optimal trading rule (OTR) algorithm
 - for, 173–174, 176–177, 192
 - strategy risk and, 211
 - triple-barrier labeling method for, 45–46, 47*f*
- Storytelling, 162
- Strategists, 7
- Strategy risk, 211–218
 - asymmetric payouts and, 213–216
 - calculating, 217, 218
 - implied betting frequency and, 215–216, 216*f*
 - implied precision and, 214–215, 215*f*
 - investment strategies and
 - understanding of, 211
 - portfolio risk differentiated from, 217
 - probabilistic Sharpe ratio (PSR)
 - similarity to, 218
 - strategy failure probability and, 216–218
 - symmetric payouts and, 211–213, 212*f*
- Structural breaks, 249–261
 - CUSUM tests in, 250–251
 - explosiveness tests in, 249, 251–259
 - sub- and super-martingale tests in, 259–261
 - types of tests in, 249–250
- Sub- and super-martingale tests, 250, 259–261
- Supernova research, 337–338, 338*f*
- Support vector machines (SVMs), 38, 101
- Supremum augmented Dickey-Fuller (SADF) test, 252–259, 253*f*, 257*f*
 - conditional ADF, 256, 256*f*, 257*f*
 - implementation of, 258–259
 - quantile ADF, 255–256
- Survivorship bias, 152
- SymPy Live, 214
- Synthetic data
 - backtesting using, 169–192

- Synthetic data (*Continued*)
 - experimental results using simulation combinations with, 176–191
 - optimal trading rule (OTR) framework using, 173–176
- Tick bars, 26–27
- Tick imbalance bars (TIBs), 29–30
- Tick rule, 282
- Tick runs bars (TRBs), 31
- Time bars
 - description of, 26
 - fixed-time horizon labeling method using, 43–44
- Time-decay factors, and sample weights, 70–71, 72*f*
- Time period, in backtesting, 196
- Time series
 - fractional differentiation applied to, 76
 - integer transformation for stationarity in, 76
 - stationarity vs. memory loss dilemma in, 75–76
- Time under water (TuW)
 - definition of, 201
 - deriving, 201
 - example of, 202*f*
 - run measurements using, 202
- Time-weighted average price (TWAP), 24, 294
- Time-weighted rate of returns (TWRR), 198–199
- Trading rules
 - investment strategies and algorithms in, 169–170
 - optimal trading rule (OTR) framework for, 173–176
 - overfitting in, 171–172
- Transaction costs, in quantitative investing, 152
- Tree clustering approaches, in asset allocation, 224–229, 225*f*, 228*f*, 232*f*
- Triple-barrier labeling method, 45–46, 47*f*, 145
- Turnover costs, 202–203
- Variance
 - boosting to reduce, 100
 - causes of, 93
 - ensemble methods to reduce, 94
 - random forest (RF) method for, 97–98
- Vectorization, 303–304
- Volume bars, 27, 44
- Volume imbalance bars (VIBs), 30–31
- Volume runs bars (VRBs), 31–32
- Volume-synchronized probability of informed trading (VPIN), 276, 282, 292
- Walk-forward (WF) method
 - backtesting using, 161–162
 - overfitting in, 155, 162
 - pitfalls of, 162
 - Sharpe ratio estimation in, 166
 - two key advantages of, 161–162
- Walk-forward timefolds method, 155
- Weighted Kendall's tau, 120–121
- Weights. *See* Class weights; Sample weights