Statistically Sound Machine Learning for Algorithmic Trading of Financial Instruments

Developing Predictive-Model-Based Trading Systems Using *TSSB*

David Aronson

with

Timothy Masters, Ph.D.
Technical Advisor

Edition 1.20

Table of Contents

Introduction 1
Two Approaches to Automated Trading 1
Predictive Modeling 2
Indicators and Targets 3
Converting Predictions to Trade Decisions 4
Testing the Trading System 5
Walkforward Testing 7
Cross Validation 9
Overlap Considerations 10
Performance Criteria 11
Model Performance Versus Financial Performance 12
Financial Relevance and Generalizability 12
Performance Statistics in *TSSB* 13
Desirable Program Features 16

A Simple Standalone Trading System 19
The Script File 19
The Audit Log 23
A Walkforward Fold 26
Out-of-Sample Results for This Fold 31
The Walkforward Summary 33

A Simple Filter System 35
The Trade File 35
The Script File 37
The Audit Log 41
Out-of-Sample Results for This Fold 43
The Walkforward Summary 45

Common Initial Commands 47
Market Price Histories and Variables 47
Quick Reference to Initial Commands 48
Detailed Descriptions 50
INTRADAY BY MINUTE 50
INTRADAY BY SECOND 50
MARKET DATE FORMAT YYMMDD 51

MARKET DATE FORMAT M_D_YYYY 51
MARKET DATE FORMAT AUTOMATIC 51
REMOVE ZERO VOLUME 52
READ MARKET LIST 52
READ MARKET HISTORIES 53
MARKET SCAN 54
RETAIN YEARS 55
RETAIN MOD 56
CLEAN RAW DATA 59
INDEX 59
READ VARIABLE LIST 60
OUTLIER SCAN 61
DESCRIBE 62
CROSS MARKET AD 63
CROSS MARKET KL 63
CROSS MARKET IQ 64
STATIONARITY 65
A Final Example 68

Reading and Writing Databases 73
Quick Reference to Database Commands 73
Detailed Descriptions 74
RETAIN MARKET LIST 74
VARIABLE IS TEXT 75
WRITE DATABASE 75
READ DATABASE 76
READ UNORDERED DATABASE 77
APPEND DATABASE 78
IS PROFIT 79
A Saving/Restoring Example 80

Creating Variables 83
Overview and Basic Syntax 83
Index Markets and Derived Variables 84
An Example of IS INDEX and MINUS INDEX 85
Multiple Indices 86
Historical Adjustment to Improve Stationarity 87
Centering 87
Scaling 88
Normalization 89
An Example of Centering, Scaling, and Normalization 89

Cross-Market Normalization 92
Pooled Variables 93
MEDIAN pooling 94
CLUMP60 Pooling 94
Mahalanobis Distance 96
Absorption Ratio 97
Trend Indicators 99
MA DIFFERENCE ShortLength LongLength Lag 99
LINEAR PER ATR HistLength ATRlength 99
QUADRATIC PER ATR HistLength ATRlength 100
CUBIC PER ATR HistLength ATRlength 100
RSI HistLength 101
STOCHASTIC K HistLength 101
STOCHASTIC D HistLength 101
PRICE MOMENTUM HistLength StdDevLength 101
ADX HistLength 102
MIN ADX HistLength MinLength 102
RESIDUAL MIN ADX HistLength MinLength 102
MAX ADX HistLength MaxLength 103
RESIDUAL MAX ADX HistLength MaxLength 103
DELTA ADX HistLength DeltaLength 103
ACCEL ADX HistLength DeltaLength 104
INTRADAY INTENSITY HistLength 104
DELTA INTRADAY INTENSITY HistLength DeltaLength 104
REACTIVITY HistLength 105
DELTA REACTIVITY HistLength DeltaDist 107
MIN REACTIVITY HistLength Dist 107
MAX REACTIVITY HistLength Dist 107
Trend-Like Indicators 108
CLOSE TO CLOSE 108
N DAY HIGH HistLength 108
N DAY LOW HistLength 109
Deviations from Trend 110
CLOSE MINUS MOVING AVERAGE HistLen ATRlen . . . 110
LINEAR DEVIATION HistLength 110
QUADRATIC DEVIATION HistLength 111
CUBIC DEVIATION HistLength 111
DETRENDED RSI DetrendedLength DetrenderLength Lookback 112

Volatility Indicators 113
ABS PRICE CHANGE OSCILLATOR ShortLen Multiplier . 113
PRICE VARIANCE RATIO HistLength Multiplier 114
MIN PRICE VARIANCE RATIO HistLen Mult Mlength 114
CHANGE VARIANCE RATIO HistLength Multiplier 114
MIN CHANGE VARIANCE RATIO HistLen Mult Mlen 115
ATR RATIO HistLength Multiplier 115
DELTA PRICE VARIANCE RATIO HistLength Multiplier 115
DELTA CHANGE VARIANCE RATIO HistLength Multiplier 116
DELTA ATR RATIO HistLength Multiplier 116
BOLLINGER WIDTH HistLength 116
DELTA BOLLINGER WIDTH HistLength DeltaLength . . . 117
N DAY NARROWER HistLength 117
N DAY WIDER HistLength 118
Indicators Involving Indices 119
INDEX CORRELATION HistLength 119
DELTA INDEX CORRELATION HistLength DeltaLength . . 119
DEVIATION FROM INDEX FIT HistLength MovAvgLength 120
PURIFIED INDEX Norm HistLen Npred Nfam Nlooks Look1 121
Basic Price Distribution Statistics 123
PRICE SKEWNESS HistLength Multiplier 123
CHANGE SKEWNESS HistLength Multiplier 123
PRICE KURTOSIS HistLength Multiplier 123
CHANGE KURTOSIS HistLength Multiplier 124
DELTA PRICE SKEWNESS HistLen Multiplier DeltaLen . . . 124
DELTA CHANGE SKEWNESS HistLen Multiplier DeltaLen 124
DELTA PRICE KURTOSIS HistLen Multiplier DeltaLen 124
DELTA CHANGE KURTOSIS HistLen Multiplier DeltaLen . . 124
Indicators That Significantly Involve Volume 125
VOLUME MOMENTUM HistLength Multiplier 125
DELTA VOLUME MOMENTUM HistLen Multiplier DeltaLen 125
VOLUME WEIGHTED MA OVER MA HistLength 126
DIFF VOLUME WEIGHTED MA OVER MA ShortDist LongDist 126
PRICE VOLUME FIT HistLength 127

DIFF PRICE VOLUME FIT ShortDist LongDist 127
DELTA PRICE VOLUME FIT HistLength DeltaDist 127
ON BALANCE VOLUME HistLength 128
DELTA ON BALANCE VOLUME HistLength DeltaDist . 128
POSITIVE VOLUME INDICATOR HistLength 129
DELTA POSITIVE VOLUME INDICATOR HistLen DeltaDist 129
NEGATIVE VOLUME INDICATOR HistLength 129
DELTA NEGATIVE VOLUME INDICATOR HistLen DeltaDist 129
PRODUCT PRICE VOLUME HistLength 130
SUM PRICE VOLUME HistLength 131
DELTA PRODUCT PRICE VOLUME HistLen DeltaDist . 131
DELTA SUM PRICE VOLUME HistLen DeltaDist 131
Entropy and Mutual Information Indicators 132
PRICE ENTROPY WordLength 133
VOLUME ENTROPY WordLength 133
PRICE MUTUAL INFORMATION WordLength 133
VOLUME MUTUAL INFORMATION WordLength 134
Indicators Based on Wavelets 135
REAL MORLET Period 137
REAL DIFF MORLET Period 138
REAL PRODUCT MORLET Period 138
IMAG MORLET Period 139
IMAG DIFF MORLET Period 139
IMAG PRODUCT MORLET Period 140
PHASE MORLET Period 140
DAUB MEAN HistLength Level 141
DAUB MIN HistLength Level 141
DAUB MAX HistLength Level 141
DAUB STD HistLength Level 141
DAUB ENERGY HistLength Level 141
DAUB NL ENERGY HistLength Level 142
DAUB CURVE HistLength Level 142
Follow-Through-Index (FTI) Indicators 143
Low-Pass Filtering and FTI Computation 144
Block Size and Channels 144
Essential Parameters for FTI calculation 144
Computing FTI 147
Automated Choice of Filter Period 148
Trends Within Trends 148

FTI Indicators Available in *TSSB* 150
FTI LOWPASS BlockSize HalfLength Period 150
FTI MINOR LOWPASS BlockSize HalfLength LowPeriod HighPeriod 151
FTI MAJOR LOWPASS BlockSize HalfLength LowPeriod HighPeriod 151
FTI FTI BlockSize HalfLength Period 151
FTI LARGEST FTI BlockSize HalfLength LowPeriod HighPeriod 152
FTI MINOR FTI BlockSize HalfLength LowPeriod HighPeriod 152
FTI MAJOR FTI BlockSize HalfLength LowPeriod HighPeriod 152
FTI LARGEST PERIOD BlockSize HalfLength LowPeriod HighPeriod 153
FTI MINOR PERIOD BlockSize HalfLength LowPeriod HighPeriod 153
FTI MAJOR PERIOD BlockSize HalfLength LowPeriod HighPeriod 153
FTI CRAT BlockSize HalfLength LowPeriod HighPeriod . . 154
FTI MINOR BEST CRAT BlockSize HalfLength LowPeriod HighPeriod 154
FTI MAJOR BEST CRAT BlockSize HalfLength LowPeriod HighPeriod 155
FTI BOTH BEST CRAT BlockSize HalfLength LowPeriod HighPeriod 155
Target Variables 156
NEXT DAY LOG RATIO 156
NEXT DAY ATR RETURN Distance 157
SUBSEQUENT DAY ATR RETURN Lead Distance 157
NEXT MONTH ATR RETURN Distance 157
HIT OR MISS Up Down Cutoff ATRdist 158
FUTURE SLOPE Ahead ATRdist 158
RSQ FUTURE SLOPE Ahead ATRdist 159

Screening Variables 161
Chi-Square Tests 162
Options for the Chi-Square Test 163
Output of the Chi-Square Test 164
Running Chi-Square Tests from the Menu 165
Nonredundant Predictor Screening 167

Options for Nonredundant Predictor Screening 170
Running Nonredundant Predictor Screening from the Menu .. 171
Examples of Nonredundant Predictor Screening 173

Models 1: Fundamentals 179
Overview and Basic Syntax 181
Mandatory Specifications Common to All Models 183
The INPUT list 183
The OUTPUT Specifier 184
Number of Inputs Chosen by Stepwise Selection 184
The Criterion to be Optimized in Indicator Selection 185
A Lower Limit on the Number or Fraction of Trades 189
Summary of Mandatory Specifications for All Models 190
Optional Specifications Common to All Models 191
Mitigating Outliers 191
Testing Multiple Stepwise Indicator Sets 192
Stepwise Indicator Selection With Cross Validation 193
When the Target Does Not Measure Profit 194
Multiple-Market Trades Based on Ranked Predictions 195
Restricting Models to Long or Short Trades 196
Prescreening For Specialist Models 196
Building a Committee with Exclusion Groups 197
Building a Committee with Resampling and Subsampling 199
Avoiding Overlap Bias 200
A Popularity Contest for Indicators 202
Bootstrap Statistical Significance Tests for Performance 204
Monte-Carlo Permutation Tests 206
An Example Using Most Model Specifications 207
Sequential Prediction 208

Models 2: The Models 213
Linear Regression 213
The MODEL CRITERION Specification for LINREG Models 213
The Identity Model 215
Quadratic Regression 216
The General Regression Neural Network 219
The Multiple-Layer Feedforward Network 220
The Number of Neurons in the First Hidden Layer 220
The Number of Neurons in the Second Hidden Layer 221
Functional Form of the Output Neuron 221

The Domain of the Neurons 222
A Basic MLFN Suitable for Most Applications 223
A Complex-Domain MLFN 224
The Basic Tree Model 225
A Forest of Trees 227
Boosted Trees 229
Operation String Models 231
Use of Constants in Operation Strings 233
Split Linear Models for Regime Regression 236
An Ordinary SPLIT LINEAR Model 238
The NOISE Version of the SPLIT LINEAR Model 239

Committees 241
Model Specifications Used by Committees 242
The AVERAGE Committee 244
The LINREG (Linear Regression) Committee 245
Constrained Linear Regression Committee 246
Models as Committees 247
Creating Component Models for Committees 248
Exclusion Groups 249
Explicit Specification of Different Indicators 250
Using Different Selection Criteria 250
Varying the Training Set by Subsampling 251
Varying the Training Set by Resampling 252

Oracles 253
Model Specifications Used by Oracles 254
Traditional Operation of the Oracle 255
Prescreen Operation of the Oracle 257
The HONOR PRESCREEN Option 257
The PRESCREEN ONLY Option 257
An Example of Prescreen Operation 259
More Complex Oracles 260

Testing Methods 263
Performance for the Entire Dataset 264
Walkforward Testing 265
Cross Validation by Time Period 266
Cross Validation using a Control Variable 267
Cross Validation by Random Blocks 269
Preserving Predictions for Trade Simulation 270

Market States as Trade Triggers 271
An Example of Simple Triggering 272
Triggering Based on State Change 275
Triggering Versus Prescreening 277
Commands Common to All Four Examples 278
Example 1: Model Specialization via PRESCREEN 280
Example 2: Unguided Specialization 285
Example 3: Triggering on High Volatility 290
Example 4: Triggering on Low Volatility 295

Permutation Training 299
The Components of Performance 302
Permutation Training and Selection Bias 306
Multiple-Market Considerations 311

Transforms 313
Expression Transforms 315
Quantities That May Be Referenced 316
Vector Operations in Expression Transforms 320
Vector-to-Scalar Functions 321
An Example with the @SIGN_AGE Function 321
Logical Evaluation in Expression Transforms 323
An Example with Logical Expressions 324
A More Complex Example 325
Principal Component Transforms 326
Invoking the Principal Components Transform 327
Tables Printed 328
An Example 329
Linear and Quadratic Regression Transforms 332
A Regression Transform Example 333
The Nominal Mapping Transform 336
Inputs and the Target 336
Gates 337
Focusing on Extreme Targets 338
Declaring the Transform and its Options 338
A Nominal Mapping Example 341
The ARMA Transform 345
The PURIFY Transform 353
Defining the Purified and Purifier Series 353
Specifying the Predictor Functions 355
Miscellaneous Specifications 356

Usage Considerations 357
A Simple Example 359

Complex Prediction Systems 363
Stacking Models and Committees 365

Graphics 371
Series Plot 372
Series + Market 374
Histogram 375
Thresholded Histogram 377
Density Map 380
Bivariate and Trivariate Plots 386
Trivariate Plots 391
Equity 393
Prediction Map 397
Indicator-Target Relationship 400
Isolating Predictability of Direction Versus Magnitude 405

Finding Independent Predictors 407
A FIND GROUPS Demonstration 410

Market Regression Classes 417
REGRESSION CLASS Demonstrations 419
The Hierarchical Method 419
The Sequential Method 423
The Leung Method 427

Developing a Stand-Alone System 431
Choosing Predictor Candidates and the Target 431
Choosing the Target 431
Quality Does Not Equal Quantity for Predictors 434
Predictor and Target Selection for this Study 436
Stationarity 439
The Problem of Outliers 440
Cross-Market Compatibility 442
Data Snooping: Friend or Foe? 444
Checking Stability with Subsampling 445
How Long Does the Model Hold Up? 449
Finding Models for a Committee 451
The Trading System 453

The Final Test 456

Trade Simulation and Portfolios 459
Writing Equity Curves 461
Performance Measures 465
Portfolios (File-Based Version) 466
A Portfolio Example 470

Integrated Portfolios 479
A FIXED Portfolio Example 482
An OOS Portfolio Example 483