Index

abnormal market conditions 320–321
abnormal returns 147–244
  attention-based buying 173–210
  news analytics relation 149–172
  news flow impact 211–230
  sentiment reversals 231–244
absolute returns 276, 279
accounting-related news 214–215, 221, 223
accuracy, algorithms 62–63
acquisitions 131–132, 145, 215, 303
active portfolio managers 19–20
Adaptive Market Hypothesis (AMH) 20
adjacency matrices 60
adjective–adverb phrase classifier 12, 57
ADRs (American depository receipts) 179
advertising 178
aggregation 130–131, 133, 153–158
  performance 64–67
  risk management framework 317
  sentiment and 63, 302–303
AGG_SCORE 303
agriculture study 84–85
Al see artificial intelligence
algorithms 45–62
  accuracy 62–63
  classification 50–52, 54–59, 68–69
  community detection 61–62, 67–68
  confusion matrix 62
  data mining 69
  dictionary-based 57–58, 216–217
  Media Sentiment® system 123
  real-time risks/returns 79–82
  trading strategies 307–310
aliasing 324
alpha-generating signals 24, 149, 153,
  309–310
alpha-generating strategies 20
alphas for buying 204–206
ambiguity filters 58–59
American depository receipts (ADRs) 179
American Stock Exchange (ASE) 181–182,
  195, 199–200, 204–205
AMERITRADE system 123–127
AMH see Adaptive Market Hypothesis
analyst revisions 221–224
  see also revisions …
  analyst role 211–212, 225
  analytic data 312
  announcements 249–250, 275, 292
  see also earnings announcements;
  information
  annualized units, risk 248
  anticipatory behavior 249
  APARCH models 256–258, 264
  Apple iPhone 115
  Arbitrage Pricing Theory (APT) 19
  ARCH models 271, 292
  article novelty 8
  artificial intelligence (AI) 69, 232
  ASE see American Stock Exchange
  Asian crisis 258
  assessment of risk 247–254
  asset management 248, 315–317
  asset-pricing predictions 175, 201–206
  asset price volatility 19–20, 22
  asset returns 211–230, 289–290
  asset-specific factor models 290–292,
  295–296
  ASX Index see S&P/ASX 200 Index
  asymmetry see volatility asymmetry
  asynchronous (unexpected) news 2, 5–6, 130,
  292
  asynchronous trading 248
  attention-based buying 173–210
  attention-grabbing stocks 20–21, 173–210
  AUD/USD exchange rate returns 84, 85
  autocorrelations 276, 279
  automated news analysis 311–314
  automation 2–3, 18
backtesting 119–121, 226–227
‘bad’ news see negative …
balanced search trees 82
bank quotes 89
  see also quote data
banking risk management 315–316
Barber, Brad M. 173–210
Barclays Capital et al. v. Theflyonthewall.com
  Inc. 113, 116–117
BARRA Inc. factor model 290, 292
base indices 80–81
Bayes classifier 50–52, 216, 301
Bayesian Classifiers 12–13
Bayesian news inclusion framework 252–253
bearish investors 197, 242–243
behavioral biases 18, 20–21, 212
behavioral finance 150
announcements 249
attention-based buying 173–210
earnings revisions 212
NORM initiative 319–322
portfolio simulation 164
private investments 262
sentiment reversals 243
BEKK models 276, 283
benchmark relativity 316
beta 151, 166, 168
biases
behavioral 212
negative news 260–261
survivor-biased archives 233
binary search trees 82
Black–Scholes options-pricing formulas 91
blogs 111–112
bonds 274
Bow Algorithm 52, 55, 56
Brown, Richard 307–310
browser-based systems 121, 123
‘bubbles’ 110–111
bullish investors 197, 238, 242–243
business cycle for news flows 221, 223
business size see company size factors
buy–sell imbalances 181–201
buy signals 231–244
buying behavior 173–210

CAD/USD exchange rate 77
calibration
factor models 291
scores 78–79
Capital Asset Pricing Model (CAPM) 19
capitalization stocks 162–163, 193, 195, 197
CAR see cumulative abnormal return
Catlin, Jeff 323–325
centrality 59–61
Chen, M.Y. 55–58
circuit breakers 23, 307–308
Clark, P.K. 272
classification
algorithms 50–52, 54–59, 68–69
errors 63
of news 214–215, 313–314
by month 218
sentiment analytics 301–303
by year 218
collective investor misbehavior 20
corporate news 144–145, 227–228
see also company . . . ; news flows
correlations
absolute returns 276
autocorrelations 276, 279
corporate news 227–228
extreme sentiment days 166, 169
implied volatility 91
metrics 64–67
serial correlation 227–228, 248, 252
short-horizon risk forecasts 248
volatility asymmetry 260, 268
covariances, risk management 151
‘crashes’ 110–111
crawlers 46–49
Credit Crisis, 2008 255, 317
CRS see Company Relevance Score
CRSP see Center for Research in Security Prices
cumulative abnormal return (CAR) 7, 8
cumulative returns 264–266
currencies see foreign exchange
current news 7
CVAR (conditional Value at Risk) 247
Das, Sanjiv R. 43–71
data analysis 4–9, 10, 155, 158–162, 320–321
data mining 69, 162
data pre-processing 89–91
data sources 4–6
  attention-based buying 176, 179–181
  firm-specific news 271, 274–276
  industry sentiment 136–137, 144–145
  market sentiment 131–132, 143–144
  news flow 213–219, 224, 228
  NORM initiative 321
  quantity vs. quality 325
  real-time risks/returns 75–77
  trading signals 224, 228
  volatility asymmetry 260
  see also news data; quote data
data structures 82, 153
datasets 213–219, 224, 228, 321
day measurements 156
decision making 117–118
defensive trading strategies 307, 311
delta 133, 139
demand–supply balance 110–111, 117
density functions 84
deseasonalized squared log returns 16, 17, 83
diBartolomeo, Dan 247–254, 289–304
  315–317
dictionaries 10, 11, 57–58, 216–217
differentiation of news 169
directional industry strategy 140–142
disagreement measure 63–64
disagreement measures 13
disambiguation of words 324–325
disclosure process 112, 113, 115
discriminant-based classification 12, 55–57
disposition effect 177
diversified managers 180–181, 190
dividend announcements 197, 198
dividend cuts 10
DJNA see Dow Jones News Analytics
document sentiment 324
Dohmen, Thomas 319–322
domain specificity, algorithms/data 45–46
dotcom technology bubble 110–111
Dow Jones News Analytics (DJNA) 232–235
Dow Jones News Service 176, 183, 232–235
downmarket conditions 268
drawdown risk 316
Duong, Huu Nhan 271–288
Dzielinski, Michal 255–270
earnings announcements 5
earnings forecasts 211–230
earnings momentum 211–212, 225, 227
earnings news releases 117, 120–123
earnings revision strategies 211–212, 221, 225
ECM see Editorials & Commentary
economic news 221
  see also financial news; macroeconomic news
economic usefulness of metrics 62, 67–68
EDA see exploratory data analysis
EDGAR see Electronic Data Gathering Analysis and Retrieval
Editorials & Commentary (ECM) 145, 303
Efficient Market Hypothesis (EMH) 19, 231, 320–321
EGARCH model 276, 281–283, 285
eigensystems 60–61
Electronic Data Gathering and Analysis and Retrieval (EDGAR) 4–5
electronic trading 261
e-mail systems 121
EMH see Efficient Market Hypothesis
empirical studies
  attention-based buying 179
  industry sentiment 138–140
  market sentiment 133–135
  news flow impact 211–230
employment announcements 274
endogeneity 182
ENS see Event Novelty Score
entity-level sentiment 324–325
entity recognition 313, 324
entry interval, Monte Carlo 240, 241
entry prices, stocks 121
EPS forecasts 221, 223
equality of variance test 88
equity analysts 19
equity coverage
  RavenPack News Analytics 30–35
  Reuters News Analytics 25–29
equity markets 274, 321
equity portfolio risk 289–304
equity price volatility, risk control/
  monitoring models 23
ES (Expected Shortfall) 320
ESS see Event Sentiment Score
Estonia 264, 268
EUR realized volatility 90–91
EURO STOXX 50 302
event analysis 82–83, 92
event-driven (unexpected) news 130, 292
event-driven trading strategies 224–227
event indices 73–108
  implied volatilities 90–91
  NewsScope Event Indices 90–91
  risk control/monitoring models 23
event novelty 129
  scores 132–133, 135–136, 144
  sentiment indices 130, 134–135
  thresholds 153–154, 156
Event Novelty Score (ENS) 132–133, 135–136, 144
event risk 73–108
Event Sentiment Score (ESS) 132–133, 135
  136, 144
event studies
  examples 83–85
  investment signals 149
  long-period analysis 151–152
  news analytics improvements 153–155
  portfolio comparison 163
  September 2008 92
  volatility asymmetry 263–264
Event Study Explorer tool 158–159, 162
event-driven asynchronous announcements 2, 5–6, 130, 292
Excel spreadsheets 47
excess returns 237–239, 241–242
exchange rates see foreign exchange
exchange-traded options 201
exit prices, stocks 121
exotic securities 317
  expansion of economy 221
expected news 130
expected news see synchronous news
Expected Shortfall (ES) 320
expected volatility 91
Expert Consensus Model 301
exploratory data analysis (EDA) 155, 158–162
  extreme returns 175, 176, 205–206
  extreme sentiment days 163–169
factor models 233, 250–251, 253, 289–297, 299
  application 18, 19–20
  factor realizations 290
  factual content, message boards 48
  false positive metrics 63
  Fama–French factor model 290–291
  fat tails issue 316
  FD (Full Disclosure) 112, 113, 118
  filtering 58–59, 130, 132, 149, 158–162
  financial market risk 247–254, 320
  see also market risk assessment
  financial news 2
  content-level 323
regulation 116
revision clusters 223–224
semantic analysis 320
  see also asynchronous news; synchronous news
financial paradigms 319–322
financial stocks 297–300
  see also stock ...
firm capitalization 162–163
Fisher’s Discriminant 55–56
fit-for-purpose systems 325
forecast errors 212
forecasting 211–230, 248, 250–251
foreign exchange (FX)
  data sources 76–77
  event analysis 82–83
  event study example 83–85
  implied volatility 89–92
  qualitative textual news 16–17
  risk control/monitoring models 23
t-statistics 86–87
foreign exchange quote data 100–102
foreign exchange–related (FXR) events 90
four-factor alphas 204–205
Friday’s news 219
Full Disclosure (FD) 112, 113, 118
fund management 18–21
fundamental models 290–292
fundamental research 24
future earnings estimates 113–114
future portfolio research 170–171
future realized volatility 91
futures returns
  prediction 129–131
  rolling over contracts 275
  volatility 271–288
FX see foreign exchange
gammas 257–258, 259
GARCH models 22
GARCH process 248–250, 256–257,
  see also EGARCH model
General Inquirer program 151, 232
generalized autoregressive conditional
  heteroskedasticity see GARCH process
generalized regression models 69
Gervais, Simon 178–179
goals, NORM initiative 321
Gonzalez, Armando 311–314
‘good’ news see positive news
goodness-of-fit test 89
Google 261–264, 269
  see also websites
Google Trends 262–264
regulation 116
revision clusters 223–224
semantic analysis 320
  see also asynchronous news; synchronous news
financial paradigms 319–322
financial stocks 297–300
  see also stock ...
firm capitalization 162–163
Fisher’s Discriminant 55–56
fit-for-purpose systems 325
forecast errors 212
forecasting 211–230, 248, 250–251
foreign exchange (FX)
  data sources 76–77
  event analysis 82–83
  event study example 83–85
  implied volatility 89–92
  qualitative textual news 16–17
  risk control/monitoring models 23
t-statistics 86–87
foreign exchange quote data 100–102
foreign exchange–related (FXR) events 90
four-factor alphas 204–205
Friday’s news 219
Full Disclosure (FD) 112, 113, 118
fund management 18–21
fundamental models 290–292
fundamental research 24
future earnings estimates 113–114
future portfolio research 170–171
future realized volatility 91
futures returns
  prediction 129–131
  rolling over contracts 275
  volatility 271–288
FX see foreign exchange
gammas 257–258, 259
GARCH models 22
GARCH process 248–250, 256–257,
  see also EGARCH model
General Inquirer program 151, 232
generalized autoregressive conditional
  heteroskedasticity see GARCH process
generalized regression models 69
Gervais, Simon 178–179
goals, NORM initiative 321
Gonzalez, Armando 311–314
‘good’ news see positive news
goodness-of-fit test 89
Google 261–264, 269
  see also websites
Google Trends 262–264
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>grammar graphs, phase lag metrics</td>
<td>67–68</td>
</tr>
<tr>
<td>Grullon, G.</td>
<td>178–179</td>
</tr>
<tr>
<td>guidance-related news</td>
<td>223, 224</td>
</tr>
<tr>
<td>Hafez, Peter Ager</td>
<td>129–146</td>
</tr>
<tr>
<td>‘hard’ news</td>
<td>214, 225</td>
</tr>
<tr>
<td>‘have to’ trades</td>
<td>249</td>
</tr>
<tr>
<td>headlines</td>
<td>14</td>
</tr>
<tr>
<td>headlines, machine learning</td>
<td>171</td>
</tr>
<tr>
<td>Healy, Alexander D.</td>
<td>73–108</td>
</tr>
<tr>
<td>heap data structure</td>
<td>82</td>
</tr>
<tr>
<td>hedge funds</td>
<td>261, 315–316</td>
</tr>
<tr>
<td>heterogeneity of beliefs</td>
<td>176, 197, 201</td>
</tr>
<tr>
<td>heuristics</td>
<td>262</td>
</tr>
<tr>
<td>hierarchy of news flows</td>
<td>219–220</td>
</tr>
<tr>
<td>high-attention days</td>
<td>174, 176–177, 198–199, 203</td>
</tr>
<tr>
<td>high-frequency data</td>
<td>271</td>
</tr>
<tr>
<td>high-frequency trading</td>
<td>169, 248, 311, 312</td>
</tr>
<tr>
<td>historical data</td>
<td>312</td>
</tr>
<tr>
<td>attention-based buying</td>
<td>176</td>
</tr>
<tr>
<td>Media Sentiment&lt;sup&gt;K&lt;/sup&gt; system</td>
<td>120–121</td>
</tr>
<tr>
<td>semantic analysis</td>
<td>320–321</td>
</tr>
<tr>
<td>hit ratios</td>
<td>134</td>
</tr>
<tr>
<td>holding period, stocks</td>
<td>225–227, 237–238, 240</td>
</tr>
<tr>
<td>honing indices</td>
<td>81</td>
</tr>
<tr>
<td>housing market bubble</td>
<td>110–111</td>
</tr>
<tr>
<td>HTML tags</td>
<td>50</td>
</tr>
<tr>
<td>implied volatility</td>
<td>89–92, 151, 251, 264, 292–299</td>
</tr>
<tr>
<td>event indices</td>
<td>90–91</td>
</tr>
<tr>
<td>NewsScope Event Indices</td>
<td>90–91, 92</td>
</tr>
<tr>
<td>qualitative textual news</td>
<td>16–17</td>
</tr>
<tr>
<td>incomplete information, market risk</td>
<td>319–320</td>
</tr>
<tr>
<td>independence of news</td>
<td>215–216</td>
</tr>
<tr>
<td>indices</td>
<td></td>
</tr>
<tr>
<td>foreign exchange (FX) implied volatility</td>
<td>16–17, 89–92</td>
</tr>
<tr>
<td>honing tools</td>
<td>81</td>
</tr>
<tr>
<td>industry-level</td>
<td>130–131, 137–138</td>
</tr>
<tr>
<td>market-level</td>
<td>132–136</td>
</tr>
<tr>
<td>see also base indices; event indices; sentiment indices</td>
<td></td>
</tr>
<tr>
<td>individual investors</td>
<td></td>
</tr>
<tr>
<td>attention-based buying</td>
<td>173–210</td>
</tr>
<tr>
<td>volatility asymmetry</td>
<td>259–261</td>
</tr>
<tr>
<td>see also private investors</td>
<td></td>
</tr>
<tr>
<td>industry applications</td>
<td>305–325</td>
</tr>
<tr>
<td>industry-level index calculation</td>
<td>137–138</td>
</tr>
<tr>
<td>industry sentiment</td>
<td>129, 130–131, 135–142, 144–145</td>
</tr>
<tr>
<td>industry-specific factor models</td>
<td>290</td>
</tr>
<tr>
<td>industry trends, news flows</td>
<td>228</td>
</tr>
<tr>
<td>inflation</td>
<td>263–264, 267, 316</td>
</tr>
<tr>
<td>information</td>
<td></td>
</tr>
<tr>
<td>arrival of 271–288</td>
<td></td>
</tr>
<tr>
<td>efficiency</td>
<td>162–163</td>
</tr>
<tr>
<td>flow 271–272, 275</td>
<td></td>
</tr>
<tr>
<td>networks</td>
<td>69</td>
</tr>
<tr>
<td>text conversion</td>
<td>45</td>
</tr>
<tr>
<td>see also announcements</td>
<td></td>
</tr>
<tr>
<td>information flow</td>
<td>17–18</td>
</tr>
<tr>
<td>information ratio, industry sentiment</td>
<td>141</td>
</tr>
<tr>
<td>information sets</td>
<td>20</td>
</tr>
<tr>
<td>informational content</td>
<td>9</td>
</tr>
<tr>
<td>see also sentiment</td>
<td></td>
</tr>
<tr>
<td>informational content of news</td>
<td>216–217, 225</td>
</tr>
<tr>
<td>insider trading regulation</td>
<td>112–113</td>
</tr>
<tr>
<td>institutional investors</td>
<td>173–210</td>
</tr>
<tr>
<td>intelligence amplification (IA)</td>
<td>3</td>
</tr>
<tr>
<td>interactive exploratory data</td>
<td>158–162</td>
</tr>
<tr>
<td>international data</td>
<td>260</td>
</tr>
<tr>
<td>international market coverage</td>
<td>171</td>
</tr>
<tr>
<td>internet usage</td>
<td>262–263</td>
</tr>
<tr>
<td>see also Google; message boards</td>
<td></td>
</tr>
<tr>
<td>intraday applications</td>
<td>49, 271–288</td>
</tr>
<tr>
<td>intraday stock index</td>
<td>271–288</td>
</tr>
<tr>
<td>investment hypothesis</td>
<td>164</td>
</tr>
<tr>
<td>investment information sources</td>
<td>117–118</td>
</tr>
<tr>
<td>investment signals</td>
<td>149, 163–168</td>
</tr>
<tr>
<td>investment strategies</td>
<td>312</td>
</tr>
<tr>
<td>investors</td>
<td></td>
</tr>
<tr>
<td>attention-based buying</td>
<td>173–210</td>
</tr>
<tr>
<td>news flow impact</td>
<td>211–230</td>
</tr>
<tr>
<td>risk management framework</td>
<td>315–316</td>
</tr>
<tr>
<td>sentiment reversals</td>
<td>238, 242–243</td>
</tr>
<tr>
<td>volatility asymmetry</td>
<td>255–270</td>
</tr>
<tr>
<td>joint distribution, MDH model</td>
<td>272–273</td>
</tr>
<tr>
<td>jumps in GARCH models</td>
<td>257</td>
</tr>
<tr>
<td>Kalev, Petko S.</td>
<td>271–288</td>
</tr>
<tr>
<td>kernel choice, SVMs</td>
<td>53</td>
</tr>
<tr>
<td>keywords/topics list</td>
<td>78, 79</td>
</tr>
<tr>
<td>Kittrell, John</td>
<td>231–244</td>
</tr>
<tr>
<td>kurtosis</td>
<td>248</td>
</tr>
<tr>
<td>lagged trading volume</td>
<td>281, 283</td>
</tr>
<tr>
<td>Lamoureux, C.G.</td>
<td>273</td>
</tr>
<tr>
<td>Laplace’s Formula</td>
<td>51</td>
</tr>
<tr>
<td>Lastrapes, W.D.</td>
<td>273</td>
</tr>
<tr>
<td>lead–lag analysis</td>
<td>67–68</td>
</tr>
<tr>
<td>learning method types</td>
<td>68–69</td>
</tr>
<tr>
<td>Lee, C.M.C.</td>
<td>177</td>
</tr>
<tr>
<td>Leinweber, David</td>
<td>149–172</td>
</tr>
<tr>
<td>Levene’s Test</td>
<td>88</td>
</tr>
<tr>
<td>leverage effect</td>
<td>255</td>
</tr>
<tr>
<td>see also volatility asymmetry</td>
<td></td>
</tr>
<tr>
<td>leverage effects</td>
<td>22–23</td>
</tr>
<tr>
<td>Lexalytics Inc.</td>
<td>323–325</td>
</tr>
</tbody>
</table>
lexicon 12
lexicons 54
likelihood ratios (LRs) 280, 282
limit orders 190
linguistics 216–217, 231, 295
liquidity 249, 314, 316
literature reviews
  news flow 211–213
  volatility 272–274
Lo, Andrew W. 73–108
log–log regressions 261
log returns 16, 17
log returns time series 85
long-only money managers 232
long-period analysis 151–152
LRs see likelihood ratios
M&A (mergers and acquisitions) news 145, 215
machine learning (ML)
  computational linguistics 216
  filtering 158
  headlines 171
  qualitative textual news 11
  machine-readable XML news feeds 7
  macroeconomic factor models 290, 291–292
  macroeconomic news 5, 74, 263, 269, 273–274
market abuse 24
market cap minimum 236–237
market capitalization 162–163, 193, 195
market conditions 9
  alpha-generating 309–310
  downmarket 268
  semantic analysis 320–321
market high, S&P 500 136, 142
market information, portfolio risk 289–304
market-maker techniques 308
market risk assessment 319–320, 322
  see also financial market risk
market sentiment 131–135
  data 143–144
  equity portfolio risk 298
  news relationship 129–146, 295–296
market surveillance 23
market uncertainty 271–272
market volatility 255, 258–259, 261–268, 292–293
market-based measures 11
marketing budgets 115
MarketPsy 20
markets
  computational linguistics measures 217
  index calculation level 132–136, 142
  news impact 110–111
  RNSE coverage 171
volatility asymmetry 255, 258–259, 261–268
  see also stock markets
Markowitz Mean Variance Analysis 289, 316
materiality 109, 112–114, 116–117
matrices 60, 62–63
MCQ ranking, DJNA archives 234
MDH see mixture of distribution hypothesis
mean reversion 206
mean scores 55
mean testing 85, 88
mean variance portfolio models 289, 316
measuring volatility asymmetry 256–257
media, see also social media
media coverage 217, 219, 259–260
media effect 231–232
media sentiment 109–128
  see also news sentiment
Media Sentiment™ system 120–128
mergers and acquisitions (M&A) news 145, 215
Merton, R. 178–179
message ambiguity 13
message boards 47–49
metadata 14, 153, 154, 308–309, 325
Metcalf’s Law 59
metrics 62–68
  alternative metrics 41–146
  risk 247, 320
Mitra, Gautam 289–304
Mitra, Leela 289–304
mixture of distribution hypothesis (MDH) 271, 272–273, 283
ML see machine learning
models 316
  application 17–24
  buying/selling behavior 174, 201–208
  computational architecture 17–18
  equity portfolio risk 289–292, 293–299
  factor models 233, 250–251, 253, 289–297, 299
  fund management 18–21
  generalized regression 69
  industry 23–24
  information flow 17–18
  intraday stock index 285–286
  quantitative models 212
  risk control/monitoring models 22–23
  sentiment classifiers 301
  trading management 18–21
momentum
  earnings momentum 211–212, 225, 227
  price momentum 129, 134–136, 309–310
  momentum investors 174
momentum managers 180, 190
Moniz, Andy 211–230
monotonicity 238
Monte Carlo null distributions 102
Monte Carlo simulations 232–233, 239–243
monthly portfolio performance 165
Morgan Stanley Capital International (MSCI) index 257, 259
Morgan Stanley High-Tech (MSH) Index
11–14, 64–67
MSCI see Morgan Stanley Capital International index
multifactor models 20, 289–291
multipliers, stock prices 111
Munz, Marion 109–128
mutual funds 179

NA see News Analytics
naive Bayes Classifier 52
naive classifier (NC) 12, 52
NASDAQ
earnings news releases 120, 121
OMXT index 264
short-sale constraints 199–200, 204–205
sort methodology 181–182, 195
natural language 11–13
natural language processing (NLP) 44, 69, 158, 216
NC see naive classifier
negative news 9
bias 260–261
equity portfolio risk 300, 301–302
flows 219, 220, 223
framework 250
qualitative textual news 10
risk control/monitoring models 22–23
negative returns 177
negative sentiment
company relevance 130
long-period analysis 151–152
market capitalization 162–163
portfolio simulation 166
reversals 232, 235, 243
as signal 308–310
systems 118–119, 121, 123
negative serial correlation 248
NEIs see NewsScope Event Indices; Reuters NewsScope Event Indices
net buying behavior 197
net news sentiment 231–234, 238
neutral news 301–302
New York Stock Exchange (NYSE) 120–121, 181–182, 193, 195, 199–200, 204–205
news, role of 248–250
news accumulation 156
news alerts 14, 75–76, 78–79, 121–127
News Analytics (NA)
data and 131–132, 136–137
definition 43
framework/techniques 43–71, 315–317
improving with aggregation 153–158
metadata 308–309
metrics 43–71
parameters 155–157
risk assessment 253
signals 163–168
stock returns relation 149–172
technology 232–235
trading and 130
underpinnings of 312
see also Thomson Reuters News Analytics
news data 4–9, 10, 150
algorithms and 45–46
classification 6–7
pre-analysis 6–9, 10
real-time risks/returns 75–76
structures 153
types 312
news flow algorithms 23–24
news flows
alpha-generating 309
analyst revisions 221–224
asset returns impact 211–230
datasets 213–219, 224, 228
relevance 219–221
news index volatility 89–92
news intensity thresholds 153, 156
news liquidity 314
see also liquidity
news-optimized risk management 319–322
news releases
attention-based buying 175–176
earnings 117, 120–123
scrapers 47–48
spillover effects 132
news sentiment
equity portfolio risk 298
future returns prediction 129–130
net news sentiment 231–234, 238
price momentum 129
systems 118–128
value measurement 109–128
news sorts 183, 190–193
news volume 78
newspaper article timeliness 214
NewsScope Event Indices (NEIs) 73–108
NewsScope Sentiment Engine 153–158, 163–171
newswires 311
Niederhoffer, V. 74
NLP see natural language processing
noise 7–8, 111–112, 307–310
noise traders 201
non-cyclical financials 161
NORM initiative 319–322
normal distribution, price changes 272
normalized real-time news scores 15
normalized scores 78–79
novelty scores 243, 309, 314
see also event novelty
null distributions 102
NYSE see New York Stock Exchange
Odean, Terrance 173–210
offensive trading strategies 307, 309–310
oil price search theme 263–264, 266
OMXT index 264
opinions, message boards 48
optimism scores 13, 57–58
optimistic sentiment 55–56
optimizer-based portfolios 169
option-implied volatility 292–294, 296–299
option markets 274
option price volatility 91, 292–293
option trader risk 251
options, attention-grabbing 174, 179, 201
options-pricing formulas 91
order execution 308
order imbalances 193
orthogonal factors 293–294
out-of-sample results 139–140
outliers 237, 257
parsing 57
part-of-speech tagging 313
PCM see Projections by Company
performance of portfolios 165
pessimism 55–56, 57–58, 232
phase lag metrics 67–68
Plexus Group 179–181
portfolio managers 19–20
portfolio risk 248, 289–304, 316–317
portfolio simulation 149, 163–168, 239–240
portfolios
characteristics 166
construction of 164–165, 168–170, 202–205, 240
metrics 68
monthly performance 165
positive news 9
equity portfolio risk 300–302
flows 219–220, 223
framework 250
qualitative textual news 10
volatility asymmetry 260
positive returns 177
positive sentiment
company relevance 130
long-period analysis 151–152
market capitalization 162–163
portfolio simulation 166
reversals 243
as signal 308–310
systems 118–119, 121, 123
positive serial correlation 248
post-event samples 83–85, 157–158
post-trade analyses 24
posting frequency, message boards 48–49
PPI (Producers’ Price Index) 274
pragmatism 109–128
pre-analysis, news data 6–9, 10
pre-event samples 83–85, 157–158
pre-news data 4, 150
pre-processing data 89–91
previous-day returns 176
price changes
alpha-generating 309–310
attention-based buying 182
‘have to’/‘want to’ trades 249
MDH model 272–273
news flows 219–220
price momentum 129, 134–136, 309–310
prices
implied volatility 91
portfolio construction 168–169
see also stock prices
private investors 22–23, 255–270
see also individual investors
probability, Bayes Classifier 50–52
Producers’ Price Index (PPI) 274
product codes 171
products 171, 305–325
professional investors see institutional investors
‘professional’ news content 325
profit and loss calculations 121
Projections by Company (PCM) 144–145, 302
proof-of-concept application 319, 321
proxies
information flow 272
volatility asymmetry 262
psychological dictionaries 10
psychological motivation 197
public disclosure, SEC 112
public relations budgets 115
publicly traded companies 115–116, 118–119, 120–123
purchases see buying behavior
pure news signals 155, 169–170
Q-statistics 279
qualitative textual news
implied volatility 16–17
quantified metrics/time series 10–17
real-time news analytics 14–16, 17
quality vs. quantity of sources 325
Quandt–Andrews Unknown Breakpoint Test 276
quantcentration effect 312–313
quantcentration effects 20
quantextual portfolios 169, 170–171
quantification
market volatility 293, 295–296
metrics 10–17, 41–146
sentiment 233–235
quantitative investor beliefs 3
quantitative models 212
quantitative price 168–169
quantity vs. quality of sources 325
quote data 100–102, 103, 104
random event studies 88
randomized treaps 82
rank stocks 18–19
RavenPack
DJNA collaboration 232
equity coverage 30–35
factor model application 20
industry sentiment 136–137, 144–145
linguistic analytics 295
market sentiment data 143–144
quantcentration effect 313
relevant news 7–8
S&P 500 companies 132–133
sentiment indices 131
tagged stories 313–314
raw real-time news scores 15
raw scores 302–303
RCM see Reports on Corporate Actions
real-time news analytics
framework 77–82
media sentiment 109, 120–121, 123
qualitative textual news 14–16, 17
risks/returns management 73–108
real-time returns/risk 92, 99, 105
realized foreign exchange volatility 16–17
realized volatility 76–77, 90–91, 264
recession effects 221, 263–265, 269
regression analysis 14, 69, 203–204, 261
regular synchronous announcements 2, 5
regulated news 112–113, 115–116, 118
release times 8–9
relevance scores
news flow datasets 214
sentiment indices 131
sentiment reversals 234–235
Thomson Reuters 308
thresholds 153–154, 156
relevant news 7–8
Reports on Corporate Actions (RCM) 145, 303
research studies, buying 177–179
retail brokerage firms 181, 183–186, 188–193, 199–200, 204–205
returns
distribution 166, 237–238, 242
NewsScope Event Indices 92, 99, 105
predictions 201–206
private investments 264–266
qualitative textual news 11, 15, 16, 17
risk and 248–249, 264–266, 292–293, 317
sentiment indices 129–130
signals for 307
sorts 182–183, 186–193
Thomson Reuters 73–108
volatility 271–288
see also abnormal returns; asset returns; stock returns
Reuters
Money News Alert 273
NewsScope Event Indices 11, 14–17, 73–108
see also Thomson Reuters
Reuters Instrument Code (RIC) 153–157
Reuters News Analytics 25–29
Reuters NewsScope Sentiment Engine (RNSE) 153–158, 163–171
reversal events 231–244, 309–310
revisions clusters 221–224
revisions strategies, earnings 211–212, 225
RIC see Reuters Instrument Code
risk 245–304
calibration models 22–23
equity portfolio risk 289–304
factors and 233
news as assessment variable 247–254
NewsScope Event Indices 92, 99, 105
volatility asymmetry 255–270
see also uncertainty
risk assessment 247–254, 319–320, 322
risk management 151
computerized scores 311
literature review 74–75
model industry application 24
News Analytics role 312, 315–317
news-optimized 319–322
Thomson Reuters NewsScope 73–108
risk sentiment measures 18–19
RNSE see Reuters NewsScope Sentiment Engine
social networks 111–112, 119
Société Générale–quoted EUR implied
volatility 91
’soft’ news 214, 225
sort methodology, buying 181–197, 201, 206
speech tagging 313
SPI 200 Futures 274–284
spillover effects 132
spot data 100–101
squared log returns 16, 17, 83
stability of metrics 62
standalone news events 215
state variables 247–254
statistical data structures 153
statistical factor models 291–292, 293–294
statistical language 11–13
statistical tests 88
statistical validity, metrics 62
stemming words 50
stock markets
behavioral bias models 20–21
bubbles 110–111
crashes 110–111
equity portfolio risk 298
human irrationality 231
Media Sentiment system 120, 121
net news sentiment 233
news value 110–111
picking/ranking models 18–19
risk control/monitoring models 22
see also trading; US stock market
stock movement 117, 119
stock prices
company size 115
media sentiment 118, 121
movement 117, 119
news impact 109, 111, 113–114
sentiment reversals 235–238, 240
stock returns
attention-grabbing stocks 183
holding period 225–227
intraday stock index 271–288
news analytics relation 149–172
sentiment indices 64–67
see also returns
stock sorting 181–197
stock tickers 120
stock-screening tools 24
stop limit–reached alerts 123
stop loss 166
strategic news 214–215, 221, 223, 227
suboptimal trading 178
’sudden market changes’ 321
supervised learning methods 68–69
supply–demand balance 110–111, 117
support vector machines (SVMs) 52–54
survivor-biased archives 233
SVMs (support vector machines) 52–54
SWAG methodology 155
synchronous (expected) news 2, 5, 130
t-distribution 88
t-hour EUR realized volatility 91
t-statistics 85–88
NewsScope Event Indices 94–98, 102
t-tests 85, 88
tagged stories 6–7, 301
DJNA archives 233
doors alerts 14
news classification 313–314
text pre-processing 50
TD AMERITRADE system 123–127
technical indicators 117
technology 232–235, 305–325
templates for classification 313
Tetlock, Paul 151–152, 212, 232
text pre-processing 50
textual news 2, 5–6, 45, 149, 323–324
quantified metrics/time series 10–17
Theflyonthewall.com Inc. case 113, 116–117
Thomson Reuters
metadata 308–309
NewsScope Event Indices 73–108
NewsScope Sentiment Engine 153–158, 163–171
sentiment measurement 323–324
Thomson Reuters News Analytics (TRNA)
25–29, 308–309, 323–324
see also Reuters News Analytics
Thomson Reuters NewsScope Data 102, 105–106
three-by-three volume/returns partition
193–194
time horizon, risk 247–248
time series 10–17
asset pricing 203
buy–sell imbalances 182–183
event analysis 82–83
market uncertainty 271–272
phase lag metrics 67–68
volatility asymmetry 259, 261
timeliness issues 157, 213–214, 247–248
TimeStamps 14
tokens 313
topic codes, RNSE 171
topics/keywords list 78, 79
trader decision support 23
trading
  asynchronous 248
  creation of news 117
DJNA products 232
high-frequency 169, 248
  news impact 117–118, 130, 311–314
SEC regulation 112–113
  signals 224, 228
strategies
  algorithms 307–310
  design 224–227
times 156
volatility asymmetry 261, 269
see also stock markets
trading management 18–21
trading volumes
  alpha-generating 309–310
  attention-based buying 175–178, 181, 202, 204–205
MDH model 272–273
  news flows 219
S&P/ASX 200 Index 274–275, 281, 283
trailing net news sentiment 234, 238
trailing windows 133–134, 139
transaction costs 165
treap data structure 82
Treasury bonds 274
TRNA see Thomson Reuters News Analytics
U-shaped buy–sell imbalance 187, 190
  uncertainty 271–272
  see also risk . . . ; volatility
underperformance 175, 203
unexpected news 130, 292
unexpected news see asynchronous news
unforeseen exposure 308
uniqueness see article novelty
universe
  of companies 133, 137–138
  sentiment reversals 232, 235–239
  of stocks 236–237
unregulated news 112–113
unscheduled news 130, 292
unscheduled news see asynchronous news
unstructured content, Lexalytics 323–324
unsupervised learning methods 68–69
US stock market
  announcements 249–250
  Media Sentiment system 120, 121
  news analytics improvements 155
  news value 110, 113
  portfolio simulation 163–168
  volatility asymmetry 263
USD/AUD exchange rate returns 84, 87
USD/CAD exchange rate returns 77
utility–attention correlation 207
validity
  event indices 82–89
  metrics 62
Value at Risk (VaR) 247, 248, 320
value managers 180, 190
value-weighted portfolios 203–205
VaR see Value at Risk
variance equality test 88
variance of portfolios 289
VCM see Venture, Corporate, Mergers & Acquisitions
VECH model 276, 283
vector distance classifier 12, 54–55
Venture, Corporate, Mergers & Acquisitions (VCM) 145, 303
visualization 158–162
VIX contract 253
volatility
  alpha-generated 309–310
  foreign exchange rates 76–77, 83–87
  futures returns 271–288
  Google and 261–264, 269
  intraday stock index 271–288
  portfolio construction 168–169
  qualitative textual news 11, 15, 16–17
  risk and 151, 292–300
  risk control/monitoring models 22–23
  September 11th 251
  volatility asymmetry 22–23, 255–270
  volatility persistence 273, 281, 283
  volume analytics 168–169
  see also news volume
  volume partitions, buying 181–182, 193–194
  volume sorts 181–182, 183–186, 190–193
  voting classifiers 58
  Vreijling, Mark 319–322
walktrap algorithm 61
Wall Street analysts 113, 115–117, 122
‘want to’ trades 249
warrants 179
websites 46–49, 111–112, 118–119
  behavioral biases 21
  risk control/monitoring models 22–23
  weekend news 296
weekly seasonality 83
weighted portfolios 68, 203–205, 297
winsorized return distributions 237–238, 242
Wolf detectors 23, 307–308
Word & Phrase Level (WLE) 144, 302
word count classifiers 54
Wu, Guojun 178

χ² test 89
XHAL computer 43–44, 69