## Index

ACF, see autocorrelation function

additive models, see also smoother, 140–142, 269

with autocorrelation, 149, 316

backfitting algorithm, 142–144, 273–274

comparison of, 136–138, 144–146, 276–277

approximate F test, 136

quadratic forms test, 137, 145, 277

significance trace, 138

degrees of freedom, 143, 274, 276

generalised, 141

and interactions, 141

projection matrices, 142–143, 273–274

for quantiles, 164

smoothing matrices, 142, 273

variance estimation, 143, 275–276

AIC, 90, 115–116, 182, 183, 195, 325

alkalinity, Round Loch of Glenhead

see data sets

analysis of variance, 15, 56

for model comparison, 88

analysis strategy, choice of, 26

anomalies, 83, 84

ANOVA see analysis of variance

AR(1), see also process, autoregressive, 97

ARMA, see process, ARMA

assumptions

checking, see plots, residual

consequences of failure, 68–69

initial, 26

transparent, 61

autocorrelation, see also autocorrelation

function, 14, 30

coefficient, 31

Durbin–Watson test for, 66

effects of, 14

on estimation, 68, 275, 316

handling, 97–102, 119–121, 148–149, 165, 260, 290, 316–317

portmanteau test for, 182–183

autocorrelation function

partial, 175–176

sample, 31–34

computation of, 32

estimator of theoretical ACF, 35, 172

for irregularly spaced data, 35

interpretation of, 32–33

limits for, 32

and missing data, 32

and model identification, 175

and seasonality, 33, 312

and trends, 33

theoretical, 34, 172

autocovariance, 31

function, 172

autoregression, see also process, autoregressive

generalized, 230

integer-valued, 230

nonlinear, 229

auxiliary information, use in analysis, 16–18, 82
backfitting, see additive models
backshift operator, 190, 203, 208, 211
bandwidth, see also kernel 129
basis, 130, 154, 239
Fourier, 130, 154
Legendre polynomial, 292, 296
spline, 131, 341
wavelet, 154–155
Bayes estimator, 81
Bayesian inference, 17, 81–82
in spatiotemporal models, 346–347
for structural models, 219, 230
Bayesian model averaging, 188
beech (fagus sylvatica), 333–334
damage categories, 334
trends in health status, 354–357
BLUP, see prediction, best linear
unbiased
bootstrap, 120–121, 164
block, 121, 147
nonparametric, 121
for nonparametric regression, 137–138, 145
number of samples, 120
parametric, 120–121
prediction intervals, 117–118, 188
test for changepoint, 110
Box–Cox transform, 96–97
boxplots, 28–30
construction of, 28–29
effect of different group sizes, 30
outliers in, 30
purpose of, 28
Brownian motion, 113, 206
fractional, 207
CCF see cross-correlation function
censoring, 257–260
and autocorrelation, 260
and extremes, 259
handling, 258–260
limit of detection, 257
changepoints, see also discontinuity
detection, 108–111
failure of standard theory, 109, 111
implications for resource
management, 284
in variance, 111
multiple, 111
tests for, 109–110
Choleski decomposition, 99, 272
Cochrane–Orcutt method, 98–99
coefficient of determination, see \( R^2 \)
cointegration, 201
collinearity, 87, 312–314
handling, 88, 323
comprehensive R archive network, 19
conditional models, see lagged
response models
control chart, 56
correlation
Pearson coefficient of, 15, 31
spatial, see also variogram, 240,
290–291, 319–320, 325
temporal, see autocorrelation
correlogram, see also autocorrelation
function, 31
residual, 66, 182
covariates, 82
categorical, 83–84
choice of, 88–91, 144–146
correlation between, 87
lagged values of, 83
orthogonal, 87
CRAN, see comprehensive R archive
network
credible intervals, 81
cross-correlation function, 38–41,
244
interpretation of, 39
wavelets as alternative to, 159
cross-validation, see smoothing
parameter, choice of
curse of dimensionality, 140
Cusum
chart, 56
test for changepoint, 110
cycles, 106–108
approximate, 172, 202,
204
in ecological populations, 192
estimation of frequency, see
periodogram analysis
phase of, 85
test for, 107
Darling Range, 285
data quality, see also inhomogeneities;  
missing data, 25–26,  
304
recording resolution, 286
data sets  
on website, 20
alkalinity in the Round Loch of  
Glenhead, 10–12
exploratory analysis, 39–41
models for, 91–93, 102–105
Dutch wind speed, 6–7, 256
exploratory analysis, 26–27,  
29–30, 33–34, 237–238
smoothed, 43–44, 134–135
STL decomposition, 152–154
trend models, 134–135,  
138–140, 146–150, 164–166
wavelet analysis, 157–159
East Atlantic Jet, model for, 203–204
North Sea haddock, 8–10
model diagnostics, 66–67,  
78–79
periodogram analysis, 107
prediction, 10, 69, 75–76,  
188–189, 198–199
trend models, 63–65, 183–185,  
196–199
ozone in eastern England, 12–14
exploratory analysis, 37–38
model for, 222–228
degrees of freedom, 88, 115
effective, 132–133, 274,  
276
for error, in nonparametric regression, 135
residual, 65
deletion  
of model terms, 90
of observations, 77
dependence, see correlation
design matrix, 70
for penalised splines, 346
detection, limits of, see censoring
detrending, see trend, removal of
deviance, 115
analysis of, 115
quasi, 118
scaled, 115
differencing, 194
fractional, 203
as high-pass filter, 195
discontinuity detection, see also  
change points, 161–162, 284
distributions
beta-binomial, 300
exponential family, 112–113
gamma, 289
relation to $\chi^2$, 119
generalised extreme value, 251
generalised Pareto, 253
inverse Gaussian, 113
von Mises, 141
downscaling, 18
Durbin–Levinson algorithm, 175
Durbin–Watson test, 66

EAJ, see East Atlantic Jet
East Atlantic Jet, see also data sets, 201
effective sample size, 97
EMEP programme, 267–268
EOF analysis, see principal  
components analysis
errors  
in data, 25, 28
unexplained variation, 63, 98
eutrophication, 307–309
exploratory analysis, purpose of, 25–26
extrapolation, see also prediction, 6
difficulties with, 10, 17, 76
into tails of distribution, 251
model based, 17–19, 83, 117–166,  
118
necessity of, 17
scenario based, 83
extremes, 5, 43, 56, 113, 250–257
block maxima, 251–253
point process characterisation,  
256–257
return period, 251
and trends, 253
threshold exceedances, 253–255
and autocorrelation, 254–255
choice of threshold, 254
covariate-dependent, 255
trends in, 252–253,  
256
F test, 65, 88–90, 115
   approximate, 136–137, 144, 276
dependence-adjusted, 149, 276
factor analysis, 243
dynamic, 212, 324–326
   with short records, 329
stationarity requirement, 243–244
factors, 84
   levels of, 84
fast Fourier transform, 32
FFT, see fast Fourier transform
filter, see also smoother; wavelets
   band-pass, 54
circular, 51
design of, 49
effect on polynomials, 45–46,
   195–196
end effects, 41, 44, 50–51, 130, 138
high-pass, 52, 195
induced structure, see preprocessing,
dangers of
Kalman, see structural models
linear, 41, 45–54
low-pass, 49
   Spencer’s 15-point, 45
   squared gain, 49
Fisher, Ronald, 15
fitted values, 65
   matrix representation, see also hat
   matrix, 72, 143, 273
forecasting, see prediction
forest, health of, see beech (*fagus*
   *sylvatica*)
Fourier, see also basis
   coefficients, 46
   frequencies, 46, 106
   theory, 46–48
GARCH models, 95, 229
general linear process, 179
generalised estimating equations, 119,
   241, 338
generalised linear models, 62
   (in table), 95, 97, 111–120,
   229
   adjusted dependent variate, 113
dispersion parameter, 112
   estimation of, 114
   linear predictor, 111
   link function, 111
   for rainfall, 289
   variance function, 112
GEOmon project, 12
GLM, see generalised linear models
GLS, see least squares, generalised
graphs
   aspect ratio, effect of, 26
   guidelines for preparation, 27–28
haddock, North Sea, see data sets
hare, snowshoe, 192
harmonics, 85
hat matrix, 72, 128
heteroscedasticity, 316, 323–324
   effect of, 68
   handling, 94–96, 98, 118–119
   types of, 95
hierarchical models, 240
homoscedasticity, 66
hypothesis tests, see tests
   increments, 206
   indicator variables, 83
   influential observations, 76–78, 100
   handling, 79
   information criterion
      Akaike’s, see AIC
deviance, 350
inhomogeneities, 12, 26, 236
innovations, 180–182
   algorithm, 180, 203
   and prediction, 186–187
standardised, 182
integrated water supply scheme, 283
interaction, see also varying coefficient
   models, 86–87, 92, 336
   and measurement units, 86, 89
interdecadal variability, 202, 204
interventions, 108–109
invertibility, 178–179, 182, 186
Kendall’s τ , 55
kernel, 129, 270
   Gaussian, 129, 270
   periodic, 141
   tricube, 151
   variable bandwidth, 130,
LAD, see least absolute deviations
lag, 31
lagged response models, see also response, lagged, 338–339
comparisons between, 101
interpretation of, 101, 297–300, 322
latent variables, 337
least absolute deviations, 79, 163
least squares, 70–71
estimators, 70, 71
properties of, 71, 96
generalised, 99–101, 186, 241
iteratively weighted, 113, 144
matrix approach to, 70–71
and maximum likelihood, 80, 81
nonlinear, 105–106
optimality, 63, 68, 96, 99, 205
penalised, 130–131
for time series models, 177–179
versus LAD, 79
weighted, 95–96, 99
locally, 129
leverage, 77, 116
likelihood ratio statistic, 115
adjusted for dependence, 291
Ljung–Box–Pierce statistic, 183
local level model, 208
as ARIMA(0,1,1), 209
signal-to-noise ratio, 221
local linear trend model, 208
local mean, see Nadaraya–Watson estimator
Loess, see Lowess
log-likelihood, 80
diffuse, 218
logistic regression, 112, 289
loglinear model, 112
Lowess, 66, 151–152
Lynx, Canadian, 192
main effects, 86
interpretation of, 86, 89
Mann–Kendall test, see also Theil–Sen estimator, 54, 109, 259
assumptions, 56
partial, 55
seasonal, 55
map
bubble, 236
contour, 236
grey-scale, 236
marginal models, 338
marginality principle, 89
Markov chain Monte Carlo, 82, 348
matrix
design, see design matrix
idempotent, 72
identity, 71
influence, see hat matrix
information, expected, 81, 114, 201, 218
and time trends, 114–115
smoothing, 128, 272
analogue of hat matrix, 128
Toeplitz, 98
trace of, 132
maximum likelihood estimation, 15, 79–81
adjusted for dependence, 290–291
for censored data, 258
for fractionally integrated models, 203
for inhomogeneous Poisson processes, 247–248
and generalised least squares, 99, 180, 181
for generalised linear models, 113–116
large sample theory, 81, 114–116
for linear trend model, 79–80
optimality, 79, 177
restricted, 240, 323, 348
for structural models, 217–218
for time series models, 179–182
likelihood factorisation, 119–120, 182, 217
of variances, 80, 114, 240
MCMC, see Markov chain Monte Carlo
mean-shift model, 109
median regression, 79, 162, 163
method of moments, 114
min/max autocorrelation factor, 244
missing data, 12–13, 25, 152
effect of, 286–287
handling, 26–27, 32, 35–37, 54, 90–91, 215, 218
possible reasons for, 27, 30, 287
mixed effects models, 240, 339–340, 347–348
model, linearity of, 62
model-building, guidelines for, 18–19, 90, 89–310, 323
modelling
  empirical, 3
  framework, statistical, 15, 57, 61
monitoring, 6
  legislation, 309
  online, 56, 110
moving average
  process, see process, moving average
  smoother, see smoother
Nadaraya–Watson estimator, 130, 222
nested models, 88, 115, 144–145, 278
normal equations, 70, 71
normalisation, see also seasonality, 15, 83
offset, 248
origin, choice of, 69
orthogonal polynomials, 87–88
outliers, see also influential observations, 26, 29
overdifferencing, 194–195
overdispersion, 246, 249
ozone, eastern England, see data sets
parameters, 62
Pearson, Karl, 15
periodicity, see cycles
periodogram analysis, 106–107
  Fisher’s test, 107
  limitations of, 107–108
permutation tests, 121, 145
phase, see cycles
plots, see also graphs; maps
  quantile–quantile, 66
    for discrete data, 116–117
  time series, 26–27
point process, 245–250
  clustered, 249–250
  inhomogeneous Poisson, 246–248, 256
diagnostics for, 249
intensity estimation for,
  marked, 250
  Poisson, 246–248
  self-exciting, 249
  spatial, 249
population
  age structure, 199
  dynamics, 191–193, 212–213
  growth, 62, 105
posterior distribution, 81
  calculating, 81–82
Prais–Winsten transformation, 100, 180
pre-whitening, 39, 185
predator–prey model, 191–193, 244
prediction, see also extrapolation
  with ARMA models, 186–188
    difference equation form, 186
    best linear unbiased, 215
  effect of model uncertainty, 188
  error, properties of, 75, 187
  intervals, 75, 118, 187
    bootstrap-based, 117–118
    for GLMs, 117–118
  lead time, 186
  long-term, 18, 188, 195, 202, 223–224, 228
  with nonparametric models, 228
  for nonstationary processes, 195
  out-of-sample verification, 188, 189
  for stationary processes, 187–188, 204
  with structural models, 213–214
preprocessing, dangers of, 49, 54, 83, 183
principal components analysis, 241–243
  loadings, 242
  stationarity requirement, 243–244
prior distribution, see also auxiliary information, use in analysis, 81
  effect on results, 82
  noninformative, 82
  random walk, 342
  spatial, 343–346
process
  AR(1), ACF of, 98, 173, 176
  AR(2), ACF of, 173, 177
  ARIMA, 194–195
effective sample size, 195
ARMA, 174–175
  ACF of, 174–175
contaminated, 211, 226
seasonal, 175
ARMA(1,1), ACF of, 174–175, 177
ARMA(2,1), 192–193
autoregressive, 171–173
  ACF of, 173, 175
quasi-cyclical behaviour, 172, 173
  stationarity of, 172–173
CAR(1), 206–207
difference–stationary, 194
fractionally integrated, 203
  ACF of, 203
Gaussian, 240, 344
intrinsic, 36
long memory, 204–205, 207
  mechanisms for, 205
Markov, 297
moving average, see also
  invertibility, 174
  ACF of, 174
  invertibility of, 179
  PACF of, 175
multivariate, 244–245
Ornstein–Uhlenbeck, 206–207
point, see point process
stationary, 34–35
  models for, 171–175, 202–205
trend–stationary, 180
vector ARMA, 244
Wiener, see Brownian motion
quantile regression, see also extremes,
  162–166, 255
linear, 162–164
model fitting, 162–164
  nonparametric, 164
quasi-likelihood, 118
R², 64–65
random effects, 239–240, 339–340
random variables, data as values of, 4, 80
random walk, see also Brownian motion,
  193–194,
  (in figure), 200, 206, 342
bivariate, 344
second order, 342
reduced model, 88
regression, see also median regression;
  quantile regression; trend,
  linear
  cumulative, 338
dynamic, 211
for spatial data, 239–240
function, 62
  additive, 141
local linear, 129–130, 270
  and autocorrelation, 272–273
  bivariate, 272
  and ordinary least squares, 130
logistic, 112, 289
with long-memory errors, 205
multiple, 62 (in table), 82–83
multivariate, 244–245
  nonparametric, see also smoother,
  127–128
  and structural models, 219–222
principal components, 88
semiparametric, 142
spatiotemporal, 341, 340
stepwise, 90
REML, see maximum likelihood estimation, restricted
residual sum of squares, see also deviance, 72, 88, 135
dependence-adjusted, 275
residuals, 65
  absolute, 66
deviance, 116
  for generalised least squares, 100–101
  for GLMs, 116–117
partial, 142, 143
Pearson, 116
  properties of, 72–74, 140
pseudo, 135
raw, 116
squared, 66
standardised, 74, 182, 218
  multivariate, 219
  for time series models, 182
  for weighted least squares, 96
response, 82
lagged, 101–102, 119, 291–292
response, (continued)
- ordinal, 336–338
- robust estimation, 79, 151
Round Loch of Glenhead, see data sets
RSS, see residual sum of squares
sandwich estimator, 98, 290–291
scatterplot, 27, 38
seasonal adjustment, see seasonality, removal of
seasonal–trend Lowess, see STL decomposition
seasonality
- Fourier representation of, 84–85
- handling, 83–85
- removal of, 30, 52–54, 83, 313
semivariogram, see variogram
serial dependence, see autocorrelation
smoother, see also additive models
- approximate F test
dependence-adjusted, 149, 276
bias and autocorrelation, 274
bias–variance trade-off, 132, 270–271
circular, 141, 271–272, 320
Kalman, see structural models
kernel, 129
linear, 128
local linear, see regression, local linear
local polynomial, 42–43
model comparison, see additive models, comparison of
moving average, 41–42, 128
reference bands, 138, 145
spline, 130–131, 220, 341
- mixed model representation, 133, 149
- and ordinary least squares, 131
variability bands, 136, 144
smoothing parameter, choice of,
132–134, 274, 281
AIC based, 140
cross-validation, 132
effect of autocorrelation, 132, 149, 157
effect on tests, 138, 145
effective degrees of freedom, 133
informal, 134
model based, 133–134, 220, 342
smoothing, fixed-interval, see structural models, Kalman smoother
SO₂, see sulfur dioxide, emissions of
software packages
BayesX, 82
Brodgar, 310
GLIMCLIM, 291
OpenBUGS, 82
R, 19
WinBUGS, 82
spectral density, 204
spectrum, line, see also periodogram, 48
splines, 131, 141, 341
- bivariate, 343
cyclical, 141
and Kalman smoother, 220
knots, 131, 341
- number of, 133
spurious regressions, 115, 199–201
standard errors, 71
- adjusted for autocorrelation, 97–98, 148–149, 276
- adjusted for spatial dependence, 290–291
incorrect, 201
likelihood based, 81
for nonparametric regression, 135–136, 144
- prediction, 76 (in table)
of trend estimate, 76 (in table)
standardisation, see normalisation
stationarity, see also process, 202
of unexplained variation, 98, 207
STL decomposition, see also data sets;
Lowess, 54, 152, 222
stochastic differential equations, 207
structural models, 207–209
- estimation for, 217–218
- Kalman filter, 213–216
- initialising, 215–216
Kalman smoother, 216–217
- nonlinear, 230
- reduced form, 209, 211
state space representation, 209–213
- measurement equation, 210
- state vector, 209
transition equation, 210
sulfur dioxide, emissions of, 10, 267–268
survival analysis, 259
t test, 15
for regression coefficient, 63, 87–88
in time series models, 182
tensor product, 343
tests
for autocorrelation, see autocorrelation
for trend, 54–56, 63
interpreting, 16, 56–57, 68, 200, 238
likelihood ratio, see likelihood ratio statistic
post hoc, 57
power of, 55–56, 145
Theil–Sen estimator, 55, 259
trend analysis
definition of, 3
reasons for, 5–6
TRIX, 309
common trends in, 322, 326–327
multistation models for, 320–321
UK acid waters monitoring network, 10
uncertainty, assessment of, 18
underdispersion, 249
units of measurement, 89, 120
variance inflation factors, 87, 312–314
variance, residual, 64
variogram, 35–37, 207
cloud, 36
nugget, 226, 319, 340
relation to ACF, 36
residual, 207
sill, 37
varying coefficient models, 160–161, 229, 239
volatility, 95
wavelet transform
wavelet transform (continued)
discrete, 155
maximal overlap, 159
wavelets, see also data sets, Dutch
wind speed, 154–159, 161
and discontinuity detection, 157
end effects, 156, 159
extremal phase, 156
least asymmetric, 156
and long memory, 205
mother wavelet, 155
scaling coefficients, 155

scaling filter, 155
shrinkage, 157
threshold choice, 157
and trends, 157
white noise, 32, 35, 63
in continuous time, 206
line spectrum of, 48
wind speed, Dutch, see data sets
WLS, see least squares, weighted

Yule’s pendulum, 173