
STATISTICAL LEARNING FOR BIG DEPENDENT DATA

DANIEL PEÑA

Universidad Carlos III de Madrid
Madrid, Spain

RUEY S. TSAY

University of Chicago
Chicago, United States

WILEY

This first edition first published 2021
© 2021 John Wiley and Sons, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of Daniel Peña and Ruey S. Tsay to be identified as the authors of this work has been asserted in accordance with law.

Registered Office

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office

111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication Data

Names: Peña, Daniel, 1948- author. | Tsay, Ruey S., 1951- author.

Title: Statistical learning for big dependent data / Daniel Peña, Ruey S. Tsay.

Description: First edition. | Hoboken, NJ : Wiley, 2021. | Series: Wiley series in probability and statistics | Includes bibliographical references and index.

Identifiers: LCCN 2020026630 (print) | LCCN 2020026631 (ebook) | ISBN 9781119417385 (cloth) | ISBN 9781119417392 (adobe pdf) | ISBN 9781119417415 (epub) | ISBN 9781119417408 (obook)

Subjects: LCSH: Big data—Mathematics. | Time-series analysis. | Data mining—Statistical methods. | Forecasting—Statistical methods.

Classification: LCC QA76.9.B45 P45 2021 (print) | LCC QA76.9.B45 (ebook) | DDC 005.7—dc23

LC record available at <https://lcn.loc.gov/2020026630>

LC ebook record available at <https://lcn.loc.gov/2020026631>

Cover Design: Wiley

Cover Images: Colorful graph Courtesy of Daniel Peña and Ruey S. Tsay, Abstract background © duncan1890/Getty Images

Set in 10/12pt TimesTenLTStd by SPi Global, Chennai, India

CONTENTS

PREFACE	xvii
1. INTRODUCTION TO BIG DEPENDENT DATA	1
1.1 Examples of Dependent Data	2
1.2 Stochastic Processes	9
1.2.1 Scalar Processes	9
1.2.1.1 Stationarity	10
1.2.1.2 White Noise Process	12
1.2.1.3 Conditional Distribution	12
1.2.2 Vector Processes	12
1.2.2.1 Vector White Noises	15
1.2.2.2 Invertibility	15
1.3 Sample Moments of Stationary Vector Process	15
1.3.1 Sample Mean	16
1.3.2 Sample Covariance and Correlation Matrices	17
1.4 Nonstationary Processes	21
1.5 Principal Component Analysis	23
1.5.1 Discussion	26
1.5.2 Properties of the PCs	27
1.6 Effects of Serial Dependence	31

Appendix 1.A: Some Matrix Theory	34
Exercises	35
References	36
2. LINEAR UNIVARIATE TIME SERIES	37
2.1 Visualizing a Large Set of Time Series	39
2.1.1 Dynamic Plots	39
2.1.2 Static Plots	44
2.2 Stationary ARMA Models	49
2.2.1 The Autoregressive Process	50
2.2.1.1 Autocorrelation Functions	51
2.2.2 The Moving Average Process	52
2.2.3 The ARMA Process	54
2.2.4 Linear Combinations of ARMA Processes	55
2.3 Spectral Analysis of Stationary Processes	58
2.3.1 Fitting Harmonic Functions to a Time Series	58
2.3.2 The Periodogram	59
2.3.3 The Spectral Density Function and Its Estimation	61
2.4 Integrated Processes	64
2.4.1 The Random Walk Process	64
2.4.2 ARIMA Models	65
2.4.3 Seasonal ARIMA Models	67
2.4.3.1 The Airline Model	69
2.5 Structural and State Space Models	71
2.5.1 Structural Time Series Models	71
2.5.2 State-Space Models	72
2.5.3 The Kalman Filter	76
2.6 Forecasting with Linear Models	78
2.6.1 Computing Optimal Predictors	78
2.6.2 Variances of the Predictions	80
2.6.3 Measuring Predictability	81
2.7 Modeling a Set of Time Series	82
2.7.1 Data Transformation	83
2.7.2 Testing for White Noise	85
2.7.3 Determination of the Difference Order	85
2.7.4 Model Identification	87
2.8 Estimation and Information Criteria	87
2.8.1 Conditional Likelihood	87
2.8.2 On-line Estimation	88
2.8.3 Maximum Likelihood (ML) Estimation	90

2.8.4	Model Selection	91
2.8.4.1	The Akaike Information Criterion (AIC)	91
2.8.4.2	The Bayesian Information Criterion (BIC)	92
2.8.4.3	Other Criteria	92
2.8.4.4	Cross-Validation	93
2.9	Diagnostic Checking	95
2.9.1	Residual Plot	96
2.9.2	Portmanteau Test for Residual Serial Correlations	96
2.9.3	Homoscedastic Tests	97
2.9.4	Normality Tests	98
2.9.5	Checking for Deterministic Components	98
2.10	Forecasting	100
2.10.1	Out-of-Sample Forecasts	100
2.10.2	Forecasting with Model Averaging	100
2.10.3	Forecasting with Shrinkage Estimators	102
	Appendix 2.A: Difference Equations	103
	Exercises	108
	References	108
3.	ANALYSIS OF MULTIVARIATE TIME SERIES	111
3.1	Transfer Function Models	112
3.1.1	Single Input and Single Output	112
3.1.2	Multiple Inputs and Multiple Outputs	118
3.2	Vector AR Models	118
3.2.1	Impulse Response Function	120
3.2.2	Some Special Cases	121
3.2.3	Estimation	122
3.2.4	Model Building	123
3.2.5	Prediction	125
3.2.6	Forecast Error Variance Decomposition	127
3.3	Vector Moving-Average Models	135
3.3.1	Properties of VMA Models	136
3.3.2	VMA Modeling	136
3.4	Stationary VARMA Models	140
3.4.1	Are VAR Models Sufficient?	140
3.4.2	Properties of VARMA Models	141
3.4.3	Modeling VARMA Process	141
3.4.4	Use of VARMA Models	142
3.5	Unit Roots and Co-Integration	147
3.5.1	Spurious Regression	148

3.5.2	Linear Combinations of a Vector Process	148
3.5.3	Co-integration	149
3.5.4	Over-Differencing	150
3.6	Error-Correction Models	151
3.6.1	Co-integration Test	152
	Exercises	157
	References	157
4.	HANDLING HETEROGENEITY IN MANY TIME SERIES	161
4.1	Intervention Analysis	162
4.1.1	Intervention with Indicator Variables	163
4.1.2	Intervention with Step Functions	165
4.1.3	Intervention with General Exogenous Variables	166
4.1.4	Building an Intervention Model	166
4.2	Estimation of Missing Values	170
4.2.1	Univariate Interpolation	170
4.2.2	Multivariate Interpolation	172
4.3	Outliers in Vector Time Series	174
4.3.1	Multivariate Additive Outliers	175
4.3.1.1	Effects on Residuals and Estimation	176
4.3.2	Multivariate Level Shift or Structural Break	177
4.3.2.1	Effects on Residuals and Estimation	177
4.3.3	Other Types of Outliers	178
4.3.3.1	Multivariate Innovative Outliers	178
4.3.3.2	Transitory Change	179
4.3.3.3	Ramp Shift	179
4.3.4	Masking and Swamping	180
4.4	Univariate Outlier Detection	180
4.4.1	Other Procedures for Univariate Outlier Detection	183
4.4.2	New Approaches to Outlier Detection	184
4.5	Multivariate Outliers Detection	189
4.5.1	VARMA Outlier Detection	189
4.5.2	Outlier Detection by Projections	190
4.5.3	A Projection Algorithm for Outliers Detection	192
4.5.4	The Nonstationary Case	193
4.6	Robust Estimation	196
4.7	Heterogeneity for Parameter Changes	199
4.7.1	Parameter Changes in Univariate Time Series	199
4.7.2	Covariance Changes in Multivariate Time Series	200
4.7.2.1	Detecting Multiple Covariance Changes	202
4.7.2.2	LR Test	202

Appendix 4.A: Cusum Algorithms	204
4.A.1 Detecting Univariate LS	204
4.A.2 Detecting Multivariate Level Shift	204
4.A.3 Detecting Multiple Covariance Changes	206
Exercises	206
References	207
5. CLUSTERING AND CLASSIFICATION OF TIME SERIES	211
5.1 Distances and Dissimilarities	212
5.1.1 Distance Between Univariate Time Series	212
5.1.2 Dissimilarities Between Univariate Series	215
5.1.3 Dissimilarities Based on Cross-Linear Dependency	222
5.2 Hierarchical Clustering of Time Series	228
5.2.1 Criteria for Defining Distances Between Groups	228
5.2.2 The Dendrogram	229
5.2.3 Selecting the Number of Groups	229
5.2.3.1 The Height and Step Plots	229
5.2.3.2 Silhouette Statistic	230
5.2.3.3 The Gap Statistic	233
5.3 Clustering by Variables	243
5.3.1 The k -means Algorithm	244
5.3.1.1 Number of Groups	246
5.3.2 k -Medoids	250
5.3.3 Model-Based Clustering by Variables	252
5.3.3.1 Maximum Likelihood (ML) Estimation of the AR Mixture Model	253
5.3.3.2 The EM Algorithm	254
5.3.3.3 Estimation of Mixture of Multivariate Normals	256
5.3.3.4 Bayesian Estimation	257
5.3.3.5 Clustering with Structural Breaks	258
5.3.4 Clustering by Projections	259
5.4 Classification with Time Series	264
5.4.1 Classification Among a Set of Models	264
5.4.2 Checking the Classification Rule	267
5.5 Classification with Features	267
5.5.1 Linear Discriminant Function	268
5.5.2 Quadratic Classification and Admissible Functions	269
5.5.3 Logistic Regression	270
5.6 Nonparametric Classification	277
5.6.1 Nearest Neighbors	277
5.6.2 Support Vector Machines	278

5.6.2.1	Linearly Separable Problems	279
5.6.2.2	Nonlinearly Separable Problems	282
5.6.3	Density Estimation	284
5.7	Other Classification Problems and Methods	286
	Exercises	287
	References	288
6.	DYNAMIC FACTOR MODELS	291
6.1	The DFM for Stationary Series	293
6.1.1	Properties of the Covariance Matrices	295
6.1.1.1	The Exact DFM	295
6.1.1.2	The Approximate DFM	297
6.1.2	Dynamic Factor and VARMA Models	299
6.2	Fitting a Stationary DFM to Data	301
6.2.1	Principal Components (PC) Estimation	301
6.2.2	Pooled PC Estimator	303
6.2.3	Generalized PC Estimator	303
6.2.4	ML Estimation	304
6.2.5	Selecting the Number of Factors	305
6.2.5.1	Rank Testing via Canonical Correlation	306
6.2.5.2	Testing a Jump in Eigenvalues	307
6.2.5.3	Using Information Criteria	307
6.2.6	Forecasting with DFM	308
6.2.7	Alternative Formulations of the DFM	314
6.3	Generalized DFM (GDFM) for Stationary Series	315
6.3.1	Some Properties of the GDFM	316
6.3.2	GDFM and VARMA Models	317
6.4	Dynamic Principal Components	317
6.4.1	Dynamic Principal Components for Optimal Reconstruction	317
6.4.2	One-Sided DPCs	318
6.4.3	Model Selection and Forecasting	320
6.4.4	One Sided DPC and GDFM Estimation	321
6.5	DFM for Nonstationary Series	324
6.5.1	Cointegration and DFM	329
6.6	GDFM for Nonstationary Series	330
6.6.1	Estimation by Generalized DPC	330
6.7	Outliers in DFMs	333
6.7.1	Factor and Idiosyncratic Outliers	333
6.7.2	A Procedure to Find Outliers in DFM	335
6.8	DFM with Cluster Structure	336
6.8.1	Fitting DFMCS	337

6.9	Some Extensions of DFM	344
6.10	High-Dimensional Case	345
6.10.1	Sparse PCs	345
6.10.2	A Structural-FM Approach	347
6.10.3	Estimation	348
6.10.4	Selecting the Number of Common Factors	349
6.10.5	Asymptotic Properties of Loading Estimates	351
	Appendix 6.A: Some R Commands	352
	Exercises	353
	References	354
7.	FORECASTING WITH BIG DEPENDENT DATA	359
7.1	Regularized Linear Models	360
7.1.1	Properties of Lasso Estimator	362
7.1.2	Some Extensions of Lasso Regression	366
7.1.2.1	Adaptive Lasso	367
7.1.2.2	Group Lasso	367
7.1.2.3	Elastic Net	368
7.1.2.4	Fused Lasso	368
7.1.2.5	SCAD Penalty	368
7.2	Impacts of Dynamic Dependence on Lasso	377
7.3	Lasso for Dependent Data	383
7.4	Principal Component Regression and Diffusion Index	388
7.5	Partial Least Squares	392
7.6	Boosting	397
7.6.1	ℓ_2 Boosting	399
7.6.2	Choices of Weak Learner	399
7.6.3	Boosting for Classification	403
7.7	Mixed-Frequency Data and Nowcasting	404
7.7.1	Midas Regression	405
7.7.2	Nowcasting	406
7.8	Strong Serial Dependence	413
	Exercises	414
	References	414
8.	MACHINE LEARNING OF BIG DEPENDENT DATA	419
8.1	Regression Trees and Random Forests	420
8.1.1	Growing Tree	420
8.1.2	Pruning	422
8.1.3	Classification Trees	422
8.1.4	Random Forests	424

8.2	Neural Networks	427
8.2.1	Network Training	429
8.3	Deep Learning	436
8.3.1	Types of Deep Networks	436
8.3.2	Recurrent NN	437
8.3.3	Activation Functions for Deep Learning	439
8.3.4	Training Deep Networks	440
8.3.4.1	Long Short-Term Memory Model	440
8.3.4.2	Training Algorithm	441
8.4	Some Applications	442
8.4.1	The Package: keras	442
8.4.2	Dropout Layer	449
8.4.3	Application of Convolution Networks	450
8.4.4	Application of LSTM	457
8.5	Deep Generative Models	466
8.6	Reinforcement Learning	466
	Exercises	467
	References	468
9.	SPATIO-TEMPORAL DEPENDENT DATA	471
9.1	Examples and Visualization of Spatio Temporal Data	472
9.2	Spatial Processes and Data Analysis	477
9.3	Geostatistical Processes	479
9.3.1	Stationary Variogram	480
9.3.2	Examples of Semivariogram	480
9.3.3	Stationary Covariance Function	482
9.3.4	Estimation of Variogram	483
9.3.5	Testing Spatial Dependence	483
9.3.6	Kriging	484
9.3.6.1	Simple Kriging	484
9.3.6.2	Ordinary Kriging	486
9.3.6.3	Universal Kriging	487
9.4	Lattice Processes	488
9.4.1	Markov-Type Models	488
9.5	Spatial Point Processes	491
9.5.1	Second-Order Intensity	492
9.6	S-T Processes and Analysis	495
9.6.1	Basic Properties	496
9.6.2	Some Nonseparable Covariance Functions	498
9.6.3	S-T Variogram	499
9.6.4	S-T Kriging	500

9.7	Descriptive S-T Models	504
9.7.1	Random Effects with S-T Basis Functions	505
9.7.2	Random Effects with Spatial Basis Functions	506
9.7.3	Fixed Rank Kriging	507
9.7.4	Spatial Principal Component Analysis	510
9.7.5	Random Effects with Temporal Basis Functions	514
9.8	Dynamic S-T Models	519
9.8.1	Space-Time Autoregressive Moving-Average Models	520
9.8.2	S-T Component Models	521
9.8.3	S-T Factor Models	521
9.8.4	S-T HMs	522
	Appendix 9.A: Some R Packages and Commands	523
	Exercises	525
	References	525

INDEX	529
--------------	------------