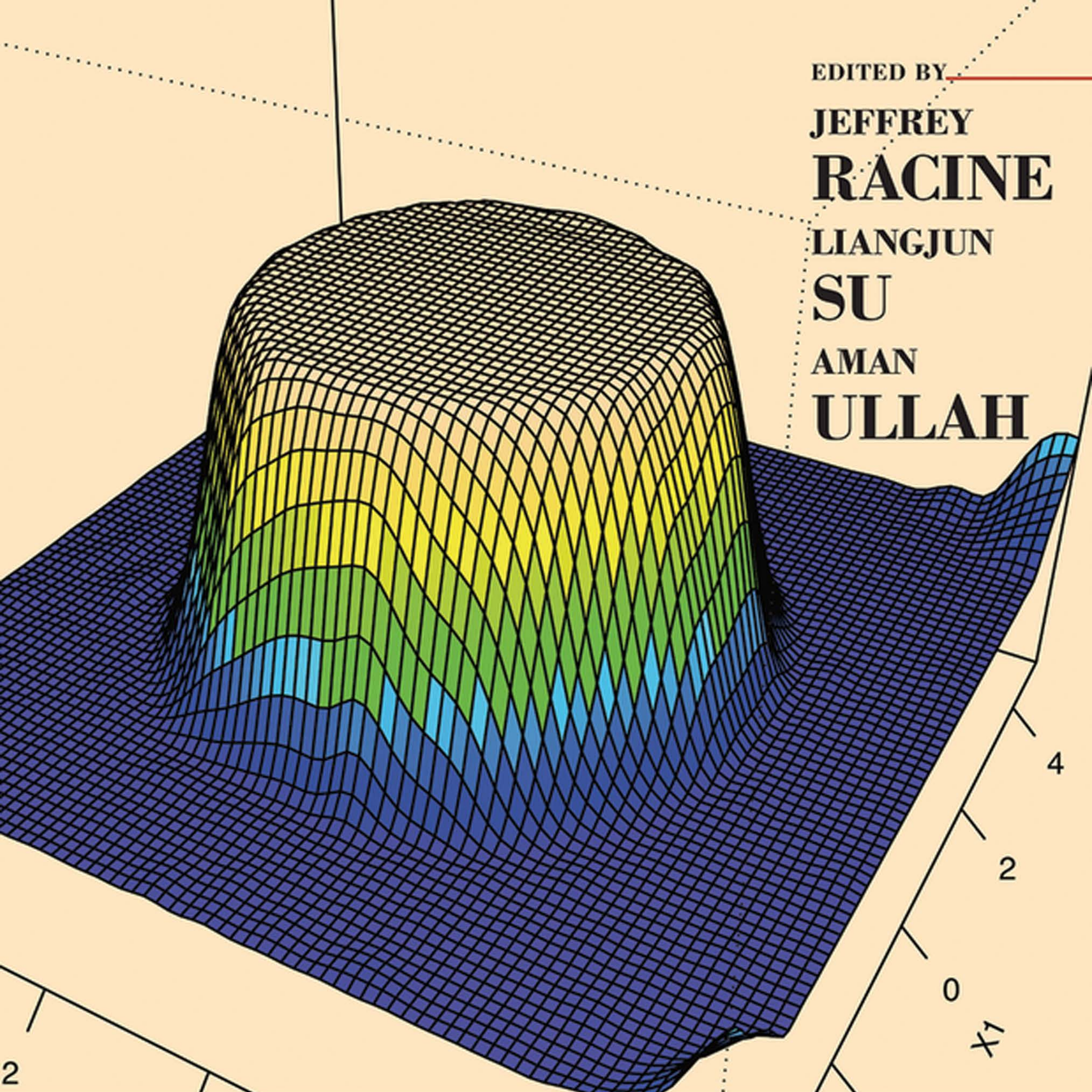
EDITED BY

JEFFREY
RACINE
LIANGJUN
SU
AMAN
ULLAH



The Oxford Handbook *of* APPLIED NONPARAMETRIC AND SEMIPARAMETRIC ECONOMETRICS AND STATISTICS

THE OXFORD HANDBOOK OF

APPLIED NONPARAMETRIC AND SEMIPARAMETRIC ECONOMETRICS AND STATISTICS

CONSULTING EDITORS

Michael Szenberg
Lubin School of Business, Pace University

Lall Ramrattan
University of California, Berkeley Extension

THE OXFORD HANDBOOK OF

APPLIED NONPARAMETRIC AND SEMIPARAMETRIC ECONOMETRICS AND STATISTICS

Edited by

JEFFREY S. RACINE, LIANGJUN SU,

and

AMAN ULLAH

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trademark of Oxford University Press
in the UK and certain other countries.

Published in the United States of America by
Oxford University Press
198 Madison Avenue, New York, NY 10016

Library of Congress Cataloging-in-Publication Data
The Oxford handbook of applied nonparametric and semiparametric econometrics and statistics / edited by
Jeffrey Racine, Liangjun Su, and Aman Ullah.
pages cm
Includes bibliographical references and index.
ISBN 978–0–19–985794–4
1. Econometrics. 2. Nonparametric statistics. I. Racine, Jeffrey Scott, 1962– editor of compilation.
II. Su, Liangjun, editor of compilation. III. Ullah, Aman, editor of compilation.
HB139.O96 2013
330.01′5195—dc23
2013002107

1 3 5 7 9 8 6 4 2
Printed in the United States of America
on acid-free paper

Contents

Preface vii
Acknowledgments xv
List of Contributors xvii

PART I METHODOLOGY

1. The Hilbert Space Theoretical Foundation of Semi-Nonparametric Modeling 3
   Herman J. Bierens
2. An Overview of the Special Regressor Method 38
   Arthur Lewbel

PART II INVERSE PROBLEMS

3. Asymptotic Normal Inference in Linear Inverse Problems 65
   Marine Carrasco, Jean-Pierre Florens, and Eric Renault
4. Identification and Well-Posedness in Nonparametric Models with Independence Conditions 97
   Victoria Zinde-Walsh

PART III ADDITIVE MODELS

5. Nonparametric Additive Models 129
   Joel L. Horowitz
6. Oracally Efficient Two-step Estimation for Additive Regression 149
   Shujie Ma and Lijian Yang
7. Additive Models: Extensions and Related Models 176
   Enno Mammen, Byeong U. Park, and Melanie Schienle

PART IV MODEL SELECTION AND AVERAGING

8. Nonparametric Sieve Regression: Least Squares, Averaging Least Squares, and Cross-Validation 215
Bruce E. Hansen

9. Variable Selection in Nonparametric and Semiparametric Regression Models 249
Liangjun Su and Yonghui Zhang

10. Data-Driven Model Evaluation: A Test for Revealed Performance 308
Jeffrey S. Racine and Christopher F. Parmeter

11. Support Vector Machines with Evolutionary Model Selection for Default Prediction 346
Wolfgang Karl Härdle, Dedy Dwi Prastyo, and Christian M. Hafner

PART V TIME SERIES

12. Series Estimation of Stochastic Processes: Recent Developments and Econometric Applications 377
Peter C. B. Phillips and Zhipeng Liao

13. Identification, Estimation, and Specification in a Class of Semilinear Time Series Models 421
Jiti Gao

14. Nonparametric and Semiparametric Estimation and Hypothesis Testing with Nonstationary Time Series 444
Yiguo Sun and Qi Li

PART VI CROSS SECTION

15. Nonparametric and Semiparametric Estimation of a Set of Regression Equations 485
Aman Ullah and Yun Wang

16. Searching for Rehabilitation in Nonparametric Regression Models with Exogenous Treatment Assignment 501
Daniel J. Henderson and Esfandiar Maasoumi

Author Index 521
Subject Index 529

Preface

Since the birth of Econometrics almost eight decades ago, theoretical and applied Econometrics and Statistics has, for the most part, proceeded along 'Classical lines which typically invokes the use of rigid user-specified parametric models, often linear. However, during the past three decades a growing awareness has emerged that results based on poorly specified parametric models could lead to misleading policy and forecasting results. In light of this, around three decades ago the subject of nonparametric Econometrics and nonparametric Statistics emerged as a field with the defining feature that models can be 'data-driven'—hence tailored to the data set at hand. Many of these approaches are described in the books by Prakasa Rao (1983), Härdle (1990), Fan and Gijbels (1996), Pagan and Ullah (1999), Yatchew (2003), Li and Racine (2007), and Horowitz (2009), and they appear in a wide range of journal outlets. The recognition of the importance of this subject along with advances in computer technology has fueled research in this area, and the literature continues to increase at an exponential rate. This pace of innovation makes it difficult for specialists and nonspecialists alike to keep abreast of recent developments. There is no single source available for those seeking an informed overview of these developments.

This handbook contains chapters that cover recent advances and major themes in the nonparametric and semiparametric domain. The chapters contained herein provide an up-to-date reference source for students and researchers who require definitive discussions of the cutting-edge developments in applied Econometrics and Statistics. Contributors have been chosen on the basis of their expertise, their international reputation, and their experience in exposing new and technical material. This handbook highlights the interface between econometric and statistical methods for nonparametric and semiparametric procedures; it is comprised of new, previously unpublished research papers/chapters by leading international econometricians and statisticians. This handbook provides a balanced viewpoint of recent developments in applied sciences with chapters covering advances in methodology, inverse problems, additive models, model selection and averaging, time series, and cross-section analysis.

Methodology

Semi-nonparametric (SNP) models are models where only a part of the model is parameterized, and the nonspecified part is an unknown function that is represented by an infinite series expansion. SNP models are, in essence, models with infinitely many parameters. In Chapter 1, Herman J. Bierens shows how orthonormal functions can be constructed along with how to construct general series representations of density and distribution functions in a SNP framework. Bierens reviews the necessary Hilbert space theory involved as well.

The term 'special regressor' originates in Lewbel (1998) and has been employed in a wide variety of limited dependent variable models including binary, ordered, and multinomial choice as well as censored regression, selection, and treatment models and truncated regression models, among others (a special regressor is an observed covariate with properties that facilitate identification and estimation of a latent variable model). In Chapter 2, Arthur Lewbel provides necessary background for understanding how and why special regressor methods work, and he details their application to identification and estimation of latent variable moments and parameters.

Inverse Problems

Ill-posed problems surface in a range of econometric models (a problem is 'well-posed' if its solution exists, is unique, and is stable, while it is 'ill-posed' if any of these conditions are violated). In Chapter 3, Marine Carrasco, Jean-Pierre Florens and Eric Renault study the estimation of a function φ in linear inverse problems of the form $T\varphi = r$, where r is only observed with error and T may be given or estimated. Four examples are relevant for Econometrics, namely, (i) density estimation, (ii) deconvolution problems, (iii) linear regression with an infinite number of possibly endogenous explanatory variables, and (iv) nonparametric instrumental variables estimation. In the first two cases T is given, whereas it is estimated in the two other cases, respectively at a parametric or nonparametric rate. This chapter reviews some main results for these models such as concepts of degree of ill-posedness, regularity of φ, regularized estimation, and the rates of convergence typically obtained. Asymptotic normality results of the regularized solution $\hat{\varphi}_\alpha$ are obtained and can be used to construct (asymptotic) tests on φ.

In Chapter 4, Victoria Zinde-Walsh provides a nonparametric analysis for several classes of models, with cases such as classical measurement error, regression with errors in variables, and other models that may be represented in a form involving convolution equations. The focus here is on conditions for existence of solutions, nonparametric identification, and well-posedness in the space of generalized functions (tempered distributions). This space provides advantages over working in function spaces by relaxing assumptions and extending the results to include a wider variety of models, for example by not requiring existence of and underlying density. Classes of (generalized) functions for which solutions exist are defined; identification conditions, partial identification, and its implications are discussed. Conditions for well-posedness are given, and the related issues of plug-in estimation and regularization are examined.

Additive Models

Additive semiparametric models are frequently adopted in applied settings to mitigate the curse of dimensionality. They have proven to be extremely popular and tend to be simpler to interpret than fully nonparametric models. In Chapter 5, Joel L. Horowitz considers estimation of nonparametric additive models. The author describes methods for estimating standard additive models along with additive models with a known or unknown link function. Tests of additivity are reviewed along with an empirical example that illustrates the use of additive models in practice.

In Chapter 6, Shujie Ma and Lijian Yang present an overview of additive regression where the models are fit by spline-backfitted kernel smoothing (SBK), and they focus on improvements relative to existing methods (i.e., Linton (1997)). The SBK estimation method has several advantages compared to most existing methods. First, as pointed out in Sperlich et al. (2002), the estimator of Linton (1997) mixed up different projections, making it uninterpretable if the real data generating process deviates from additivity, while the projections in both steps of the SBK estimator are with respect to the same measure. Second, the SBK method is computationally expedient, since the pilot spline estimator is much faster computationally than the pilot kernel estimator proposed in Linton (1997). Third, the SBK estimator is shown to be as efficient as the "oracle smoother" uniformly over any compact range, whereas Linton (1997) proved such 'oracle efficiency' only at a single point. Moreover, the regularity conditions needed by the SBK estimation procedure are natural and appealing and close to being minimal. In contrast, higher-order smoothness is needed with growing dimensionality of the regressors in Linton and Nielsen (1995). Stronger and more obscure conditions are assumed for the two-stage estimation proposed by Horowitz and Mammen (2004).

In Chapter 7, Enno Mammen, Byeong U. Park and Melanie Schienle give an overview of smooth backfitting estimators in additive models. They illustrate their wide applicability in models closely related to additive models such as (i) nonparametric regression with dependent errors where the errors can be transformed to white noise by a linear transformation, (ii) nonparametric regression with repeatedly measured data, (iii) nonparametric panels with fixed effects, (iv) simultaneous nonparametric equation models, and (v) non- and semiparametric autoregression and GARCH-models. They review extensions to varying coefficient models, additive models with missing observations, and the case of nonstationary covariates.

Model Selection and Averaging

"Sieve estimators" are a class of nonparametric estimator where model complexity increases with the sample size. In Chapter 8, Bruce Hansen considers "model selection" and "model averaging" of nonparametric sieve regression estimators. The concepts of series and sieve approximations are reviewed along with least squares estimates of sieve approximations and measurement of estimator accuracy by integrated mean-squared error (IMSE). The author demonstrates that the critical issue in applications is selection of the order of the sieve, because the IMSE greatly varies across the choice. The author adopts the cross-validation criterion as an estimator of mean-squared forecast error and IMSE. The author extends existing optimality theory by showing that cross-validation selection is asymptotically IMSE equivalent to the infeasible best sieve approximation, introduces weighted averages of sieve regression estimators, and demonstrates how averaging estimators have lower IMSE than selection estimators.

In Chapter 9, Liangjun Su and Yonghui Zhang review the literature on variable selection in nonparametric and semiparametric regression models via shrinkage. The survey includes simultaneous variable selection and estimation through the methods of least absolute shrinkage and selection operator (Lasso), smoothly clipped absolute deviation (SCAD), or their variants, with attention restricted to nonparametric and semiparametric regression models. In particular, the author considers variable selection in additive models, partially linear models, functional/varying coefficient models, single index models, general nonparametric regression models, and semiparametric/nonparametric quantile regression models.

In Chapter 10, Jeffrey S. Racine and Christopher F. Parmeter propose a data-driven approach for testing whether or not two competing approximate models are equivalent in terms of their expected true error (i.e., their expected performance on unseen data drawn from the same DGP). The test they consider is applicable in cross-sectional and time-series settings, furthermore, in time-series settings their method overcomes two of the drawbacks associated with dominant approaches, namely, their reliance on only one split of the data and the need to have a sufficiently large 'hold-out' sample for these tests to possess adequate power. They assess the finite-sample performance of the test via Monte Carlo simulation and consider a number of empirical applications that highlight the utility of the approach.

Default probability (the probability that a borrower will fail to serve its obligation) is central to the study of risk management. Bonds and other tradable debt instruments are the main source of default for most individual and institutional investors. In contrast, loans are the largest and most obvious source of default for banks. Default prediction is becoming more and more important for banks, especially in risk management, in order to measure their clients degree of risk. In Chapter 11, Wolfgang Härdle, Dedy Dwi Prastyo and Christian Hafner consider the use of Support Vector Machines (SVM) for modeling default probability. SVM is a state-of-the-art nonlinear classification technique that is well-suited to the study of default risk. This chapter emphasizes SVM-based default prediction applied to the CreditReform database. The SVM parameters are optimized by using an evolutionary algorithm (the so-called "Genetic Algorithm") and show how the "imbalanced problem" may be overcome by the use of "down-sampling" and "oversampling."

Time Series

In Chapter 12, Peter C. B. Phillips and Zhipeng Liao consider an overview of recent developments in series estimation of stochastic processes and some of their applications in Econometrics. They emphasize the idea that a stochastic process may, under certain conditions, be represented in terms of a set of orthonormal basis functions, giving a series representation that involves deterministic functions. Several applications of this series approximation method are discussed. The first shows how a continuous function can be approximated by a linear combination of Brownian motions (BMs), which is useful in the study of spurious regression. The second application utilizes the series representation of BM to investigate the effect of the presence of deterministic trends in a regression on traditional unit-root tests. The third uses basis functions in the series approximation as instrumental variables to perform efficient estimation of the parameters in cointegrated systems. The fourth application proposes alternative estimators of long-run variances in some econometric models with dependent data, thereby providing autocorrelation robust inference methods in these models. The authors review work related to these applications and ongoing research involving series approximation methods.

In Chapter 13, Jiti Gao considers some identification, estimation, and specification problems in a class of semilinear time series models. Existing studies for the stationary time series case are reviewed and discussed, and Gao also establishes some new results for the integrated time series case. The author also proposes a new estimation method and establishes a new theory for a class of semilinear nonstationary autoregressive models.

Nonparametric and semiparametric estimation and hypothesis testing methods have been intensively studied for cross-sectional independent data and weakly dependent time series data. However, many important macroeconomics and financial data are found to exhibit stochastic and/or deterministic trends, and the trends can be nonlinear in nature. While a linear model may provide a decent approximation to a nonlinear model for weakly dependent data, the linearization can result in severely biased approximation to a nonlinear model with nonstationary data. In Chapter 14, Yiguo Sun and Qi Li review some recent theoretical developments in nonparametric and semiparametric techniques applied to nonstationary or near nonstationary variables. First, this chapter reviews some of the existing works on extending the I(0), I(1), and cointegrating relation concepts defined in a linear model to a nonlinear framework, and it points out some difficulties in providing satisfactory answers to extend the concepts of I(0), I(1), and cointegration to nonlinear models with persistent time series data. Second, the chapter reviews kernel estimation and hypothesis testing for nonparametric and semiparametric autoregressive and cointegrating models to explore unknown nonlinear relations among I(1) or near I(1) process(es). The asymptotic mixed normal results of kernel estimation generally replace asymptotic normality results usually obtained for weakly dependent data. The authors also discuss kernel estimation of semiparametric varying coefficient regression models with correlated but not cointegrated data. Finally, the authors discuss the concept of co-summability introduced by Berengner-Rico and Gonzalo (2012), which provides an extension of cointegration concepts to nonlinear time series data.

Cross Section

Sets of regression equations (SREs) play a central role in Econometrics. In Chapter 15, Aman Ullah and Yun Wang review some of the recent developments for the estimation of SRE within semi- and nonparametric frameworks. Estimation procedures for various nonparametric and semiparametric SRE models are presented including those for partially linear semiparametric models, models with nonparametric autocorrelated errors, additive nonparametric models, varying coefficient models, and models with endogeneity.

In Chapter 16, Daniel J. Henderson and Esfandiar Maasoumi suggest some new directions in the analysis of nonparametric models with exogenous treatment assignment. The nonparametric approach opens the door to the examination of potentially different distributed outcomes. When combined with cross-validation, it also identifies potentially irrelevant variables and linear versus nonlinear effects. Examination of the distribution of effects requires distribution metrics, such as stochastic dominance tests for ranking based on a wide range of criterion functions, including dollar valuations. They can identify subgroups with different treatment outcomes, and they offer an empirical demonstration based on the GAIN data. In the case of one covariate (English as the primary language), there is support for a statistical inference of uniform first-order dominant treatment effects. The authors also find several others that indicate second- and higher-order dominance rankings to a statistical degree of confidence.

Jeffrey S. Racine
Liangjun Su
Aman Ullah

References

Fan, J. & Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, New York.

Horowitz, J. L. (2009), *Semiparametric and Nonparametric Methods in Econometrics*, Springer-Verlag.

Li, Q. & Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press, New York.

Prakasa Rao, B. L. S. (1983), *Nonparametric Functional Estimation*, Academic Press, Orlando, FL.

Yatchew, A. J. (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, New York.

Acknowledgments

We would like to express our deep gratitude to each of the the contributors without whom this volume would not be possible. We would also like to acknowledge the guidance and encouragement of the staff at Oxford University Press—in particular, Terry Vaughn and Cathryn Vaulman, whose tireless efforts and keen eye have smoothed the production of this handbook. Finally, we would like to thank Damaris Carlos, who helped us with many different types of tasks.

List of Contributors

Herman J. Bierens, Professor Emeritus of Economics, Pennsylvania State University

Marine Carrasco, Professor of Economics, University of Montreal

Jean-Pierre Florens, Professor of Mathematics, Toulouse School of Economics; Research Director, Institut D'Economie Industrielle

Jiti Gao, Professor, Department of Econometrics and Business Statistics, Monash University

Christian M. Hafner, Professor, School of Economics, Catholic University of Louvain

Bruce E. Hansen, Trygve Haavelmo Professor, Department of Economics, University of Wisconsin, Madison

Wolfgang Karl Härdle, Ladislaus von Bortkiewicz Chair of Statistics, Department of Economics and Business Administration, Humboldt University

Daniel J. Henderson, J. Weldon and Delores Cole Faculty Fellow of Finance and Legal, Department of Economics, University of Alabama

Joel L. Horowitz, Charles E. and Emma H. Morrison Professor, Department of Economics, Northwestern University

Arthur Lewbel, Professor, Department of Economics, Boston College

Qi Li, Hugh Roy Cullen Professor in Liberal Arts, Department of Economics, Texas A&M University

Zhipeng Liao, Assistant Professor, Department of Economics, University of California, Los Angeles

Shujie Ma, Assistant Professor, Department of Statistics, University of California, Riverside

Esfandiar Maasoumi, Arts and Sciences Distinguished Professor, Department of Economics, Emory University

Enno Mammen, Professor, Department of Economics, University of Mannheim

Byeong U. Park, Professor, Department of Statistics, Seoul National University

Christopher F. Parmeter, Assistant Professor, Department of Economics, University of Miami

Peter C. B. Phillips, Sterling Professor of Economics and Professor of Statistics, Yale University; Alumni Distinguished Professor of Economics, University of Auckland; Adjunct professor, University of Southampton and Singapore Management University; Co-Director, Centre for Financial Econometrics, Singapore Management University

Dedy Dwi Prastyo, PhD student, School of Business & Economics, Humboldt University

Jeffrey S. Racine, Senator William McMaster Chair in Econometrics and Professor of Economics and Statistics, McMaster University

Eric Renault, Professor, Department of Economics, University of Montreal for Carrasco, Toulouse School of Economics for Florens, Brown University for Renault.

Melanie Schienle, Professor, School of Economics & Management, Leibniz University Hannover

Liangjun Su, Professor, School of Economics, Singapore Management University

Yiguo Sun, Professor, Department of Economics & Finance, University of Guelph

Aman Ullah, Distinguished Professor, Department of Economics, University of California, Riverside

Yun Wang, Assistant Professor, School of International Trade and Economics, University of International Business and Economics

Lijian Yang, Professor, Department of Statistics and Probability and an Adjunct Professor for the Center for Global Change and Earth Observations, Michigan State University

Yonghui Zhang, Assistant Professor, School of Economics, Renmin University of China, China

Victoria Zinde-Walsh, Professor, Department of Economics, McGill University

PART I

METHODOLOGY

CHAPTER 1

THE HILBERT SPACE THEORETICAL FOUNDATION OF SEMI-NONPARAMETRIC MODELING

HERMAN J. BIERENS

1.1. Introduction

Semi-nonparametric (SNP) models are models where only a part of the model is parameterized, and the nonspecified part is an unknown function that is represented by an infinite series expansion. Therefore, SNP models are, in essence, models with infinitely many parameters. The parametric part of the model is often specified as a linear index, that is, a linear combination of conditioning and/or endogenous variables, with the coefficients involved the parameters of interests, which we will call the structural parameters. Although the unknown function involved is of interest as well, the parameters in its series expansion are only of interest insofar as they determine the shape of this function.

The theoretical foundation of series expansions of functions is Hilbert space theory, in particular the properties of Hilbert spaces of square integrable real functions. Loosely speaking, Hilbert spaces are vector spaces with properties similar to those of Euclidean spaces. As is well known, any vector in the Euclidean space $\mathbb{R}^k$ can be represented by a linear combination of k orthonormal vectors. Similarly, in Hilbert spaces of functions, there exist sequences of orthonormal functions such that any function in this space can be represented by a linear combination of these orthonormal functions. Such orthonormal sequences are called complete.

The main purpose of this chapter is to show how these orthonormal functions can be constructed and how to construct general series representations of density and distribution functions. Moreover, in order to explain why this can be done, I will review the necessary Hilbert space theory involved as well.

The standard approach to estimate SNP models is sieve estimation, proposed by Grenander (1981). Loosely speaking, sieve estimation is like standard parameter estimation, except that the dimension of the parameter space involved increases to infinity with the sample size. See Chen (2007) for a review of sieve estimation. However, the main focus of this chapter is on SNP modeling rather than on estimation.

Gallant (1981) was the first econometrician to propose Fourier series expansions as a way to model unknown functions. See also Eastwood and Gallant (1991) and the references therein. However, the use of Fourier series expansions to model unknown functions has been proposed earlier in the statistics literature. See, for example, Kronmal and Tarter (1968).

Gallant and Nychka (1987) consider SNP modeling and sieve estimation of Heckman's (1979) sample selection model, where the bivariate error distribution of the latent variable equations is modeled semi-nonparametrically using a bivariate Hermite polynomial expansion of the error density.

Another example of an SNP model is the mixed proportional hazard (MPH) model proposed by Lancaster (1979), which is a proportional hazard model with unobserved heterogeneity. Elbers and Ridder (1982) and Heckman and Singer (1984) have shown that under mild conditions the MPH model is nonparametrically identified. The latter authors propose to model the distribution function of the unobserved heterogeneity variable by a discrete distribution. Bierens (2008) and Bierens and Carvalho (2007) use orthonormal Legendre polynomials to model semi-nonparametrically the unobserved heterogeneity distribution of interval-censored mixed proportional hazard models and bivariate mixed proportional hazard models, respectively.

However, an issue with the single-spell MPH model is that for particular specifications of the baseline hazard, its efficiency bound is singular, which implies that any consistent estimator of the Euclidean parameters in the MPH model involved converges at a slower rate than the square root of the sample size. See Newey (1990) for a general review of efficiency bounds, and see Hahn (1994) and Ridder and Woutersen (2003) for the efficiency bound of the MPH model. On the other hand, Hahn (1994) also shows that in general the multiple-spell MPH model does not suffer from this problem, which is confirmed by the estimation results of Bierens and Carvalho (2007).

This chapter is organized as follows. In Section 1.2 I will discuss three examples of SNP models,[1] with focus on semiparametric identification. The SNP index regression model is chosen as an example because it is one of the few SNP models where the unknown function involved is not a density or distribution function. The two other examples are the bivariate MPH model in Bierens and Carvalho (2007) and the first-price auction model in Bierens and Song (2012, 2013), which have been chosen because these papers demonstrate how to do SNP modeling and estimation in practice, and in both models the unknown function involved is a distribution function. Section 1.3 reviews Hilbert space theory. In Section 1.4 it will be shown how to generate various sequences of orthonormal polynomials, along with what kind of Hilbert spaces they span. Moreover, it will also be shown how these results can be applied to the SNP index regression model. In Section 1.5 various nonpolynomial complete orthonormal sequences of functions will be derived. In Section 1.6 it will be shown how arbitrary density and distribution functions can be represented by series expansions in terms of complete orthonormal sequences of functions, along with how these results can be applied to the bivariate MPH model in Bierens and Carvalho (2007) and to the first-price auction model in Bierens and Song (2012, 2013). In Section 1.7 I will briefly discuss the sieve estimation approach, and in Section 1.8 I will make a few concluding remarks.

Throughout this chapter I will use the following notations. The well-known indicator function will be denoted by $\mathbf{1}(\cdot)$, the set of positive integers will be denoted by $\mathbb{N}$, and the set of non-negative integers, $\mathbb{N} \cup \{0\}$, by $\mathbb{N}_0$. The abbreviation "a.s." stands for "almost surely"—that is, the property involved holds with probability 1—and "a.e." stands for "almost everywhere," which means that the property involved holds except perhaps on a set with Lebesgue measure zero.

1.2. Examples of SNP Models

1.2.1. The SNP Index Regression Model

Let Y be a dependent variable satisfying $E[Y^2] < \infty$, and let $X \in \mathbb{R}^k$ be a vector of explanatory variables. As is well known, the conditional expectation $E[Y|X]$ can be written as $E[Y|X] = g_0(X)$, where $g_0(x)$ is a Borel measurable real function on $\mathbb{R}^k$.[2] Newey (1997) proposed to estimate $g_0(x)$ by sieve estimation via a multivariate series expansion. However, because there are no parameters involved, the resulting estimate of $g_0(x)$ can only be displayed and interpreted graphically, which in practice is only possible for $k \leq 2$. Moreover, to approximate a bivariate function $g_0(x)$ by a series expansion of order n requires n^2 parameters.[3] Therefore, a more practical approach is the following.

Suppose that there exists a $\beta_0 \in \mathbb{R}^k$ such that $E[Y|X] = E[Y|\beta_0'X]$ a.s. Then there exists a Borel measurable real function $f(x)$ on $\mathbb{R}$ such that $E[Y|X] = f(\beta_0'X)$ a.s. Because for any nonzero constant c, $E[Y|\beta_0'X] = E[Y|c\beta_0'X]$ a.s., identification of f requires to normalize β_0 in some way, for example by setting one component of β_0 to 1. Thus, in the case $k \geq 2$, let $X = (X_1, X_2')'$ with $X_2 \in \mathbb{R}^{k-1}$, and $\beta_0 = (1, \theta_0')'$ with $\theta_0 \in \mathbb{R}^{k-1}$, so that

$$E[Y|X] = f(X_1 + \theta_0'X_2) \qquad \text{a.s.} \tag{1.1}$$

To derive further conditions for the identification of f and θ_0, suppose that for some $\theta_* \neq \theta_0$ there exists a function f_* such that $f(X_1 + \theta_0'X_2) = f_*(X_1 + \theta_*'X_2)$ a.s. Moreover, suppose that the conditional distribution of X_1 given X_2 is absolutely continuous with support $\mathbb{R}$. Then conditional on X_2, $f(x_1 + \theta_0'X_2) = f_*(x_1 + \theta_0'X_2 + (\theta_* - \theta_0)'X_2)$ a.s.

for all $x_1 \in \mathbb{R}$. Consequently, for arbitrary $z \in \mathbb{R}$ we may choose $x_1 = z - \theta_0' X_2$, so that

$$f(z) = f_*(z + (\theta_* - \theta_0)' X_2) \qquad \text{a.s. for all } z \in \mathbb{R}. \tag{1.2}$$

If $f(z)$ is constant, then $E[Y|X] = E[Y]$ a.s., so let us exclude this case. Then (1.2) is only possible if $(\theta_* - \theta_0)' X_2$ is a.s. constant, which in turn implies that $(\theta_* - \theta_0)'(X_2 - E[X_2]) = 0$ a.s. and thus $(\theta_* - \theta_0)' E[(X_2 - E[X_2])(X_2 - E[X_2])'](\theta_* - \theta_0) = 0$. Therefore, if $\mathrm{Var}[X_2]$ is nonsingular, then $\theta_* = \theta_0$.

Summarizing, it has been shown that the following results hold.

Theorem 1.1. *The function $f(z)$ and the parameter vector θ_0 in the index regression model* (1.1) *are identified if*

(a) $\Pr[E(Y|X) = E(Y)] < 1$;
(b) *The conditional distribution of X_1 given X_2 is absolutely continuous with support $\mathbb{R}$;*
(c) *The variance matrix of X_2 is finite and nonsingular.*
Moreover, in the case $X \in \mathbb{R}$ the regression function $f(z)$ is identified for all $z \in \mathbb{R}$ if the distribution of X is absolutely continuous with support $\mathbb{R}$.

The problem how to model $f(z)$ semi-nonparametrically and how to estimate f and θ_0 will be addressed in Section 1.4.4.

1.2.2. The MPH Competing Risks Model

Consider two durations, T_1 and T_2. Suppose that conditional on a vector X of covariates and a common unobserved (heterogeneity) variable V, which is assumed to be independent of X, the durations T_1 and T_2 are independent, that is, $\Pr[T_1 \leq t_1, T_2 \leq t_2|X, V] = \Pr[T_1 \leq t_1|X, V].\Pr[T_2 \leq t_2|X, V]$. This is a common assumption in bivariate survival analysis. See van den Berg (2000). If the conditional distributions of the durations T_1 and T_2 are of the mixed proportional hazard type, then their survival functions conditional on X and V take the form $S_i(t|X, V) = \Pr[T_i > t|X, V] = \exp(-V \exp(\beta_i' X)\Lambda_i(t|\alpha_i))$, $i = 1, 2$, where $\Lambda_i(t|\alpha_i) = \int_0^t \lambda_i(\tau|\alpha_i) d\tau$, $i = 1, 2$, are the integrated baseline hazards depending on parameter vectors α_i.

This model is also known as the competing risks model. It is used in Bierens and Carvalho (2007) to model two types of recidivism durations of ex-convicts, namely (a) the time T_1 between release from prison and the first arrest for a misdemeanor and (b) the time T_2 between release from prison and the first arrest for a felony, with Weibull baseline hazards, that is,

$$\begin{aligned} &\lambda(t|\alpha_i) = \alpha_{i,1}\alpha_{i,2} t^{\alpha_{i,2}-1}, \quad \Lambda(t|\alpha_i) = \alpha_{i,1} t^{\alpha_{i,2}}, \qquad \alpha_{i,1} > 0,\ \alpha_{i,2} > 0, \\ &\text{with } \alpha_i = (\alpha_{i,1}, \alpha_{1,2})', \qquad i = 1, 2, \end{aligned} \tag{1.3}$$

where $\alpha_{i,1}$ is a scale factor.

In this recidivism case we only observe $T = \min(T_1, T_2)$ together with a discrete variable D that is 1 if $T_2 > T_1$ and 2 if $T_2 \leq T_1$. Thus, $D = 1$ corresponds to rearrest for a misdemeanor and $D = 2$ corresponds to rearrests for a felony. Then conditional on X and V, $\Pr[T > t, D = i|X, V] = \int_t^\infty V \exp(-V(\exp(\beta_1' X)\Lambda(\tau|\alpha_1) + \exp(\beta_2' X)\Lambda(\tau|\alpha_2))) \cdot \exp(\beta_i' X)\lambda(\tau|\alpha_i)\ d\tau$, $i = 1,2$, which is not hard to verify. Integrating V out now yields

$$
\begin{aligned}
&\Pr[T > t, D = i|X] \\
&= \int_t^\infty \int_0^\infty v \exp\left(-v\left(\exp\left(\beta_1' X\right)\Lambda(\tau|\alpha_1) + \exp\left(\beta_2' X\right)\Lambda(\tau|\alpha_2)\right)\right) dG(v) \\
&\qquad \times \exp\left(\beta_i' X\right)\lambda(\tau|\alpha_i)\, d\tau, \qquad i = 1,2,
\end{aligned} \tag{1.4}
$$

where $G(v)$ is the (unknown) distribution function of V.

It has been shown in Bierens and Carvalho (2006), by specializing the more general identification results of Heckman and Honore (1989) and Abbring and van den Berg (2003), that under two mild conditions the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ and the distribution function G are identified. One of these conditions is that the variance matrix of X is finite and nonsingular. The other condition is that $E[V] = 1$,[4] so that (1.4) can be written as

$$
\begin{aligned}
&\Pr[T > t, D = d|X] \\
&= \int_t^\infty H\left(\exp\left(-\left(\exp\left(\beta_1' X\right)\Lambda(\tau|\alpha_1) + \exp\left(\beta_2' X\right)\Lambda(\tau|\alpha_2)\right)\right)\right) \\
&\qquad \times \exp\left(\beta_d' X\right)\lambda(\tau|\alpha_d)\, d\tau, \qquad d = 1,2,
\end{aligned} \tag{1.5}
$$

where

$$
H(u) = \int_0^\infty v u^v\ dG(v) \tag{1.6}
$$

is a distribution function on the unit interval $[0,1]$. Thus,

Theorem 1.2. *If the variance matrix of X is finite and nonsingular, then the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ and the distribution function $H(u)$ in the MPH competing risks Weibull model* (1.5) *are identified.*

Proof. (Bierens and Carvalho, 2006, 2007). ■

It follows now straightforwardly from (1.5) that, given a random sample $\{T_j, D_j, X_j\}_{j=1}^N$ from (T, D, X), the log-likelihood function involved takes the form $\ln(L_N(\alpha_1, \alpha_2, \beta_1, \beta_2, H)) = \sum_{j=1}^N \ell(T_j, D_j, X_j|\alpha_1, \alpha_2, \beta_1, \beta_2, H)$, where

$$
\begin{aligned}
&\ell(T, D, X|\alpha_1, \alpha_2, \beta_1, \beta_2, H) \\
&= \ln\left(H\left(\exp\left(-\left(\exp\left(\beta_1' X\right)\Lambda(T|\alpha_1) + \exp\left(\beta_2' X\right)\Lambda(T|\alpha_2)\right)\right)\right)\right) \\
&\qquad + (2 - D)\left(\beta_1' X + \ln(\lambda(T|\alpha_1))\right) + (D - 1)\left(\beta_2' X + \ln(\lambda(T|\alpha_2))\right).
\end{aligned} \tag{1.7}
$$

At this point the distribution function $H(u)$ representing the distribution of the unobserved heterogeneity is treated as a parameter. The problem of how to model $H(u)$ semi-nonparametrically will be addressed in Section 1.6.

Note that the duration $T = \min(T_1, T_2)$ in Bierens and Carvalho (2007) is only observed over a period $[0, \overline{T}]$, where $\overline{T}$ varies only slightly per ex-inmate, so that T is right-censored. Therefore, the actual log-likelihood in Bierens and Carvalho (2007) is more complicated than displayed in (1.7).

1.2.3. First-Price Auctions

A first price-sealed bids auction (henceforth called *first-price auction*) is an auction with $I \geq 2$ potential bidders, where the potential bidder's values for the item to be auctioned off are independent and private, and the bidders are symmetric and risk neutral. The reservation price p_0, if any, is announced in advance and the number I of potential bidders is known to each potential bidder.

As is well known, the equilibrium bid function of a first-price auction takes the form

$$\beta(v|F,I) = v - \frac{1}{F(v)^{I-1}} \int_{p_0}^{v} F(x)^{I-1}\,dx \qquad \text{for } v > p_0 > \underline{v}, \tag{1.8}$$

if the reservation price p_0 is binding, and

$$\beta(v|F,I) = v - \frac{1}{F(v)^{I-1}} \int_{0}^{v} F(x)^{I-1}\,dx \qquad \text{for } v > \underline{v}, \tag{1.9}$$

if the reservation price p_0 is nonbinding, where $F(v)$ is the value distribution, $I \geq 2$ is the number of potential bidders, and $\underline{v} \geq 0$ is the lower bound of the support of $F(v)$. See, for example, Riley and Samuelson (1981) or Krishna (2002). Thus, if the reservation price p_0 is binding, then, with V_j the value for bidder j for the item to be auctioned off, this potential bidder issues a bid $B_j = \beta(V_j|F,I)$ according to bid function (1.8) if $V_j > p_0$ and does not issue a bid if $V_j \leq p_0$, whereas if the reservation price p_0 is not binding, each potential bidder j issues a bid $B_j = \beta(V_j|F,I)$ according to bid function (1.9). In the first-price auction model the individual values V_j, $j = 1, \ldots, I$, are assumed to be independent random drawing from the value distribution F. The latter is known to each potential bidder j, and so is the number of potential bidders, I.

Guerre et al. (2000) have shown that if the value distribution $F(v)$ is absolutely continuous with density $f(v)$ and bounded support $[\underline{v}, \overline{v}]$, $\overline{v} < \infty$, then $f(v)$ is nonparametrically identified from the distribution of the bids. In particular, if the reservation price is nonbinding, then the inverse bid function is $v = b + (I-1)^{-1}\Lambda(b)/\lambda(b)$, where v is a private value, b is the corresponding bid, and $\Lambda(b)$ is the distribution function of the bids with density $\lambda(b)$. Guerre et al. (2000) propose to estimate the latter two functions via nonparametric kernel methods, as $\hat{\Lambda}(b)$ and $\hat{\lambda}(b)$, respectively. Using the pseudo-private values $\widetilde{V} = B + (I-1)^{-1}\hat{\Lambda}(B)/\hat{\lambda}(B)$, where each B is an observed bid, the density $f(v)$ of the private value distribution can now be estimated by kernel density estimation.

Bierens and Song (2012) have shown that the first-price auction model is also nonparametrically identified if instead of the bounded support condition, the value distribution F in (1.8) and (1.9) is absolutely continuous on $(0,\infty)$ with connected support[5] and finite expectation. As an alternative to the two-step nonparametric approach of Guerre et al. (2000), Bierens and Song (2012) propose a simulated method of moments sieve estimation approach to estimate the true value distribution $F_0(v)$, as follows. For each SNP candidate value distribution F, generate simulated bids according to the bid functions (1.8) or (1.9) and then minimize the integrated squared difference of the empirical characteristic functions of the actual bids and the simulated bids to the SNP candidate value distributions involved.

This approach has been extended in Bierens and Song (2013) to first-price auctions with auction-specific observed heterogeneity. In particular, given a vector X of auction-specific covariates, Bierens and Song (2013) assume that $\ln(V) = \theta' X + \varepsilon$, where X and ε are independent. Denoting the distribution function of $\exp(\varepsilon)$ by F, the conditional distribution of V given X then takes the form $F(v \exp(-\theta' X))$.

1.3. Hilbert Spaces

1.3.1. Inner Products

As is well known, in a Euclidean space $\mathbb{R}^k$ the inner product of a pair of vectors $x = (x_1,\ldots,x_k)'$ and $y = (y_1,\ldots,y_k)'$ is defined as $x'y = \sum_{m=1}^k x_m y_m$, which is a mapping $\mathbb{R}^k \times \mathbb{R}^k \to \mathbb{R}$ satisfying $x'y = y'x$, $(cx)'y = c(x'y)$ for arbitrary $c \in \mathbb{R}$, $(x+y)'z = x'z + y'z$, and $x'x > 0$ if and only if $x \neq 0$. Moreover, the norm of a vector $x \in \mathbb{R}^k$ is defined as $||x|| = \sqrt{x'x}$, with associated metric $||x-y||$. Of course, in $\mathbb{R}$ the inner product is the ordinary product $x \cdot y$.

Mimicking these properties of inner product, we can define more general inner products with associated norms and metrics as follows.

Definition 1.1. *An inner product on a real vector space $\mathcal{V}$ is a real function $\langle x,y\rangle$: $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$ such that for all x, y, z in $\mathcal{V}$ and all c in $\mathbb{R}$, we obtain the following:*

1. $\langle x,y\rangle = \langle y,x\rangle$.
2. $\langle cx,y\rangle = c\langle x,y\rangle$.
3. $\langle x+y,z\rangle = \langle x,z\rangle + \langle y,z\rangle$.
4. $\langle x,x\rangle > 0$ *if and only if* $x \neq 0$.
   Given an inner product, the associated norm and metric are defined as $||x|| = \sqrt{\langle x,x\rangle}$ *and* $||x-y||$, *respectively.*

As is well known from linear algebra, for vectors $x, y \in \mathbb{R}^k$, $|x'y| \leq ||x||.||y||$, which is known as the Cauchy–Schwarz inequality. This inequality carries straightforwardly over to general inner products:

Theorem 1.3. *(Cauchy–Schwarz inequality)* $|\langle x, y\rangle| \leq ||x||.||y||$.

1.3.2. Convergence of Cauchy Sequences

Another well-known property of a Euclidean space is that every Cauchy sequence has a limit in the Euclidean space involved.[6] Recall that a sequence of elements x_n of a metric space with metric $||x - y||$ is called a Cauchy sequence if $\lim_{\min(k,m)\to\infty} ||x_k - x_m|| = 0$.

Definition 1.2. *A Hilbert space $\mathcal{H}$ is a vector space endowed with an inner product and associated norm and metric such that every Cauchy sequence has a limit in $\mathcal{H}$.*

Thus, a Euclidean space is a Hilbert space, but Hilbert spaces are much more general than Euclidean spaces.

To demonstrate the role of the Cauchy convergence property, consider the vector space $C[0,1]$ of continuous real functions on $[0,1]$. Endow this space with the inner product $\langle f, g\rangle = \int_0^1 f(u)g(u)\,du$ and associated norm $||f|| = \sqrt{\langle f, f\rangle}$ and metric $||f - g||$. Now consider the following sequence of functions in $C[0,1]$:

$$f_n(u) = \begin{cases} 0 & \text{for} \quad 0 \leq u < 0.5, \\ 2^n(u - 0.5) & \text{for} \quad 0.5 \leq u < 0.5 + 2^{-n}, \\ 1 & \text{for} \quad 0.5 + 2^{-n} \leq u \leq 1, \end{cases}$$

for $n \in \mathbb{N}$. It is an easy calculus exercise to verify that f_n is a Cauchy sequence in $C[0,1]$. Moreover, it follows from the bounded convergence theorem that $\lim_{n\to\infty} ||f_n - f|| = 0$, where $f(u) = \mathbf{1}(u > 0.5)$. However, this limit $f(u)$ is discontinuous in $u = 0.5$, and thus $f \notin C[0,1]$. Therefore, the space $C[0,1]$ is not a Hilbert space.

1.3.3. Hilbert Spaces Spanned by a Sequence

Let $\mathcal{H}$ be a Hilbert space and let $\{x_k\}_{k=1}^{\infty}$ be a sequence of elements of $\mathcal{H}$. Denote by

$$\mathcal{M}_m = \text{span}(\{x_j\}_{j=1}^{m})$$

the subspace spanned by $x_1, \ldots, x_m$; that is, $\mathcal{M}_m$ consists of all linear combinations of $x_1, \ldots, x_m$. Because every Cauchy sequence in $\mathcal{M}_m$ takes the form $z_n = \sum_{i=1}^m c_{i,n} x_i$, where the $c_{i,n}$'s are Cauchy sequences in $\mathbb{R}$ with limits $c_i = \lim_{n\to\infty} c_{i,n}$, it follows trivially that $\lim_{n\to\infty} ||z_n - z|| = 0$, where $z = \sum_{i=1}^m c_i x_i \in \mathcal{M}_m$. Thus, $\mathcal{M}_m$ is a Hilbert space.

Definition 1.3. *The space* $\mathcal{M}_\infty = \overline{\cup_{m=1}^{\infty}\mathcal{M}_m}$ [7] *is called the space spanned by* $\{x_j\}_{j=1}^{\infty}$, *which is also denoted by span*$(\{x_j\}_{j=1}^{\infty})$.

Let x_n be a Cauchy sequence in $\mathcal{M}_\infty$. Then x_n has a limit $\overline{x} \in \mathcal{H}$, that is, $\lim_{n\to\infty} ||x_n - \overline{x}|| = 0$. Suppose that $\overline{x} \notin \mathcal{M}_\infty$. Because $\mathcal{M}_\infty$ is closed, there exists an $\varepsilon > 0$ such that the set $\mathcal{N}(\overline{x}, \varepsilon) = \{x \in \mathcal{H} : ||x - \overline{x}|| < \varepsilon\}$ is completely outside $\mathcal{M}_\infty$, that is, $\mathcal{N}(\overline{x}, \varepsilon) \cap \mathcal{M}_\infty = \emptyset$. But $\lim_{n\to\infty} ||x_n - \overline{x}|| = 0$ implies that there exists an $\underline{n}(\varepsilon)$ such that $x_n \in \mathcal{N}(\overline{x}, \varepsilon)$ for all $n > \underline{n}(\varepsilon)$, hence $x_n \notin \mathcal{M}_\infty$ for all $n > \underline{n}(\varepsilon)$, which contradicts $x_n \in \mathcal{M}_\infty$ for all n. Thus,

Theorem 1.4. $\mathcal{M}_\infty$ *is a Hilbert space.*

In general, $\mathcal{M}_\infty$ is smaller than $\mathcal{H}$, but as we will see there exist Hilbert spaces $\mathcal{H}$ containing a sequence $\{x_j\}_{j=1}^{\infty}$ for which $\mathcal{M}_\infty = \mathcal{H}$. Such a sequence is called complete:

Definition 1.4. *A sequence* $\{x_k\}_{k=1}^{\infty}$ *in a Hilbert space* $\mathcal{H}$ *is called complete if* $\mathcal{H} =$ *span*$(\{x_j\}_{j=1}^{\infty})$.

Of particular importance for SNP modeling are Hilbert spaces spanned by a complete orthonormal sequence, because in that case the following approximation result holds.

Theorem 1.5. *Let* $\{x_j\}_{j=1}^{\infty}$ *a complete orthonormal sequence in a Hilbert space* $\mathcal{H}$, *that is,* $\langle x_i, x_j\rangle = 1(i = j)$ *and* $\mathcal{H} = \mathrm{span}(\{x_j\}_{j=1}^{\infty})$. *For an arbitrary* $y \in \mathcal{H}$, *let* $\widehat{y}_n = \sum_{j=1}^{n}\langle y, x_j\rangle x_j$. *Then* $\lim_{n\to\infty} ||y - \widehat{y}_n|| = 0$ *and* $\sum_{j=1}^{\infty}\langle y, x_j\rangle^2 = ||y||^2$.

This result is a corollary of the fundamental projection theorem:

Theorem 1.6. *Let* $\mathcal{S}$ *be a sub-Hilbert space of a Hilbert space* $\mathcal{H}$. *Then for any* $y \in \mathcal{H}$ *there exists a* $\widehat{y} \in \mathcal{S}$ *(called the projection of* y *on* $\mathcal{S}$*) such that* $||y - \widehat{y}|| = \inf_{z\in\mathcal{S}} ||y - z||$. *Moreover, the projection residual* $u = y - \widehat{y}$ *satisfies* $\langle u, z\rangle = 0$ *for all* $z \in \mathcal{S}$. [8]

Now observe that $\widehat{y}_n$ in Theorem 1.5 is the projection of y on $\mathcal{M}_n = \mathrm{span}(\{x_j\}_{j=1}^{n})$, with residual $u_n = y - \widehat{y}_n$ satisfying $\langle u_n, y_n\rangle = 0$ for all $y_n \in \mathcal{M}_n$, and that due to $y \in \mathrm{span}(\{x_j\}_{j=1}^{\infty}) = \overline{\cup_{m=1}^{\infty}\mathcal{M}_m}$ there exists a sequence $y_n \in \mathcal{M}_n$ such that $\lim_{n\to\infty} ||y - y_n|| = 0$. Then $||y - \widehat{y}_n||^2 = \langle u_n, y - \widehat{y}_n\rangle = \langle u_n, y\rangle = \langle u_n, y - y_n\rangle \leq ||u_n||.||y - y_n|| \leq ||y||.||y - y_n|| \to 0$, where the first inequality follows from the Cauchy–Schwarz inequality while the second inequality follows from the fact that $||u_n||^2 \leq ||y||^2$. Moreover, the result $\sum_{j=1}^{\infty}\langle y, x_j\rangle^2 = ||y||^2$ in Theorem 1.5 follows from the fact that $||y||^2 = \langle y, y\rangle = \lim_{n\to\infty}\langle \widehat{y}_n, y\rangle = \lim_{n\to\infty}\sum_{j=1}^{n}\langle y, x_j\rangle^2$.

1.3.4. Examples of Non-Euclidean Hilbert Spaces

Consider the space $\mathcal{R}$ of random variables defined on a common probability space $\{\Omega, \mathcal{F}, P\}$ with finite second moments, endowed with the inner product $\langle X, Y\rangle =$

$E[X.Y]$ and associated norm $||X|| = \sqrt{\langle X, X\rangle} = \sqrt{E[X^2]}$ and metric $||X - Y||$. Then we have the following theorem.

Theorem 1.7. *The space $\mathcal{R}$ is a Hilbert space.*[9]

This result is the basis for the famous Wold (1938) decomposition theorem, which in turn is the basis for time series analysis.

In the rest of this chapter the following function spaces play a key role.

Definition 1.5. *Given a probability density $w(x)$ on $\mathbb{R}$, the space $L^2(w)$ is the space of Borel measurable real functions f on $\mathbb{R}$ satisfying $\int_{-\infty}^{\infty} f(x)^2 w(x)\, dx < \infty$, endowed with the inner product $\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g(x)w(x)\, dx$ and associated norm $||f|| = \sqrt{\langle f, f\rangle}$ and metric $||f - g||$. Moreover, $L^2(a, b)$, $-\infty \leq a < b \leq \infty$, is the space of Borel measurable real functions on (a, b) satisfying $\int_a^b f(x)^2\, dx$, with inner product $\langle f, g\rangle = \int_a^b f(x)g(x)\, dx$ and associated norm and metric.*

Then for $f, g \in L^2(w)$, we have $\langle f, g\rangle = E[f(X)g(X)]$, where X is a random drawing from the distribution with density $w(x)$; hence from Theorem 1.7 we obtain the following theorem.

Theorem 1.8. *The space $L^2(w)$ is a Hilbert space.*

Also $L^2(a, b)$ is a Hilbert space, as will be shown in Section 1.5.

In general the result $\lim_{n\to\infty} ||y - \widehat{y}_n|| = 0$ in Theorem 1.5 does not imply that $\lim_{n\to\infty} \widehat{y}_n = y$, as the latter limit may not be defined, and even if so, $\lim_{n\to\infty} \widehat{y}_n$ may not be equal to y. However, in the case $\mathcal{H} = L^2(w)$ the result $\lim_{n\to\infty} ||y - \widehat{y}_n|| = 0$ implies $\lim_{n\to\infty} \widehat{y}_n = y$, in the following sense.

Theorem 1.9. *Let $\{\rho_m(x)\}_{m=0}^{\infty}$ be a complete orthonormal sequence in $L^2(w)$,*[10] *and let X be a random drawing from the density w. Then for every function $f \in L^2(w)$, $\Pr[f(X) = \lim_{n\to\infty} \sum_{m=0}^{n} \gamma_m \rho_m(X)] = 1$, where $\gamma_m = \int_{-\infty}^{\infty} \rho_m(x) f(x) w(x)\, dx$ with $\sum_{m=0}^{\infty} \gamma_m^2 = \int_{-\infty}^{\infty} f(x)^2 w(x)\, dx$.*

Proof. Denote $f_n(x) = \sum_{m=0}^{n} \gamma_m \rho_m(x)$, and recall from Theorem 1.5 that $\sum_{m=0}^{\infty} \gamma_m^2 = ||f||^2 < \infty$. It follows now that

$$E[(f(X) - f_n(X))^2] = \int_{-\infty}^{\infty} (f(x) - f_n(x))^2 w(x)\, dx = \sum_{m=n+1}^{\infty} \gamma_m^2 \to 0$$

as $n \to \infty$; hence by Chebyshev's inequality, $p\lim_{n\to\infty} f_n(X) = f(X)$. As is well known,[11] the latter is equivalent to the statement that for every subsequence of n there exists a further subsequence m_k, for example, such that $\Pr[\lim_{k\to\infty} f_{m_k}(X) = f(X)] = 1$, and the same applies to any further subsequence m_{k_n} of m_k:

$$\Pr\left[\lim_{n\to\infty} f_{m_{k_n}}(X) = f(X)\right] = 1. \tag{1.10}$$

Given n, there exists a natural number k_n such that $m_{k_n-1} < n \leq m_{k_n}$, and for such a k_n we obtain

$$E\left[\left(f_{m_{k_n}}(X) - f_n(X)\right)^2\right] = E\left[\left(\sum_{j=n+1}^{m_{k_n}} \gamma_m \rho_m(X)\right)^2\right] = \sum_{j=n+1}^{m_{k_n}} \gamma_m^2 \leq \sum_{j=m_{k_n-1}+1}^{m_{k_n}} \gamma_m^2,$$

hence

$$\sum_{n=0}^{\infty} E[(f_{m_{k_n}}(X) - f_n(X))^2] \leq \sum_{n=0}^{\infty} \sum_{j=m_{k_n-1}+1}^{m_{k_n}} \gamma_m^2 \leq \sum_{n=0}^{\infty} \gamma_n^2 < \infty.$$

By Chebyshev's inequality and the Borel–Cantelli lemma,[12] the latter implies

$$\Pr\left[\lim_{n\to\infty} (f_{m_{k_n}}(X) - f_n(X))\right] = 1. \tag{1.11}$$

Combining (1.10) and (1.11), the theorem follows. ■

1.4. Orthonormal Polynomials and the Hilbert Spaces They Span

1.4.1. Orthonormal Polynomials

Let $w(x)$ be a density function on $\mathbb{R}$ satisfying

$$\int_{-\infty}^{\infty} |x|^k \, w(x) \, dx < \infty \qquad \text{for all } k \in \mathbb{N}, \tag{1.12}$$

and let $p_k(x|w)$ be a sequence of polynomials in $x \in \mathbb{R}$ of order $k \in \mathbb{N}_0$ such that $\int_{-\infty}^{\infty} p_k(x|w)p_m(x|w)w(x)\ dx = 0$ if $k \neq m$. In words, the polynomials $p_k(x|w)$ are *orthogonal* with respect to the density function $w(x)$. These orthogonal polynomials can be generated recursively by the three-term recurrence relation (hereafter referred to as TTRR)

$$p_{k+1}(x|w) + (b_k - x)\, p_k(x|w) + c_k p_{k-1}(x|w) = 0, \qquad k \geq 1, \tag{1.13}$$

starting from $p_0(x|w) = 1$ and $p_1(x|w) = x - \int_0^1 z.w(z)\ dz$, for example, where

$$b_k = \frac{\int_{-\infty}^{\infty} x \cdot p_k(x|w)^2 w(x)\ dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)\ dx}, \qquad c_k = \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)\ dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)\ dx}. \tag{1.14}$$

See, for example, Hamming (1973).

Defining

$$\overline{p}_k(x|w) = \frac{p_k(x|w)}{\sqrt{\int_{-\infty}^{\infty} p_k(y|w)^2 w(y)\, dy}} \tag{1.15}$$

yields a sequence of *orthonormal* polynomials with respect to $w(x)$:

$$\int_{-\infty}^{\infty} \overline{p}_k(x|w)\overline{p}_m(x|w)w(x)\, dx = \mathbf{1}(k = m). \tag{1.16}$$

It follows straightforwardly from (1.13) and (1.15) that these orthonormal polynomials can be generated recursively by the TTRR

$$a_{k+1}.\overline{p}_{k+1}(x|w) + (b_k - x)\overline{p}_k(x|w) + a_k.\overline{p}_{k-1}(x|w) = 0, \qquad k \in \mathbb{N}, \tag{1.17}$$

starting from $\overline{p}_0(x|w) = 1$ and

$$\overline{p}_1(x|w) = \frac{x - \int_{-\infty}^{\infty} z \cdot w(z)\, dz}{\sqrt{\int_{-\infty}^{\infty} \left(y - \int_{-\infty}^{\infty} z \cdot w(z)\, dz\right)^2 w(y)\, dy}},$$

where b_k is the same as in (1.14) and

$$a_k = \frac{\sqrt{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)\, dx}}{\sqrt{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)\, dx}}.$$

The sequence is $\overline{p}_k(x|w)$ uniquely determined by $w(x)$, except for signs. In other words, $|\overline{p}_k(x|w)|$ is unique. To show this, suppose that there exists another sequence $\overline{p}_k^*(x|w)$ of orthonormal polynomials w.r.t. $w(x)$. Since $\overline{p}_k^*(x|w)$ is a polynomial of order k, we can write $\overline{p}_k^*(x|w) = \sum_{m=0}^{k} \beta_{m,k}\overline{p}_m(x|w)$. Similarly, we can write $\overline{p}_k(x|w) = \sum_{m=0}^{k} \alpha_{m,.k}\overline{p}_m^*(x|w)$. Then for $j < k$, we have

$$\int_{-\infty}^{\infty} \overline{p}_k^*(x|w)\overline{p}_j(x|w)w(x)\, dx = \sum_{m=0}^{j} \alpha_{m,j} \int_{-\infty}^{\infty} \overline{p}_k^*(x|w)\overline{p}_m^*(x|w)w(x)\, dx = 0$$

and

$$\begin{aligned}\int_{-\infty}^{\infty} \overline{p}_k^*(x|w)\overline{p}_j(x|w)w(x)\, dx &= \sum_{m=0}^{k} \beta_{m,k} \int_{-\infty}^{\infty} \overline{p}_m(x|w)\overline{p}_j(x|w)w(x)\, dx \\ &= \beta_{j,k} \int_{-\infty}^{\infty} \overline{p}_j(x|w)^2 w(x)\, dx = \beta_{j,k};\end{aligned}$$

hence $\beta_{j,.k} = 0$ for $j < k$ and thus $\overline{p}_k^*(x|w) = \beta_{k,k}\overline{p}_k(x|w)$. Moreover, by normality,

$$1 = \int_{-\infty}^{\infty} \overline{p}_k^*(x|w)^2 w(x)\, dx = \beta_{k,k}^2 \int_{-\infty}^{\infty} \overline{p}_k(x|w)^2 w(x)\, dx = \beta_{k,k}^2,$$

so that $\overline{p}_k^*(x|w) = \pm\overline{p}_k(x|w)$. Consequently, $|\overline{p}_k(x|w)|$ is unique. Thus, we have the following theorem.

Theorem 1.10. *Any density function $w(x)$ on $\mathbb{R}$ satisfying the moment conditions* (1.12) *generates a unique sequence of orthonormal polynomials, up to signs. Consequently, the sequences a_k and b_k in the TTRR* (1.17) *are unique.*

1.4.2. Examples of Orthonormal Polynomials

1.4.2.1. Hermite Polynomials

If $w(x)$ is the density of the standard normal distribution,

$$w_{\mathcal{N}[0,1]}(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi},$$

the orthonormal polynomials involved satisfy the TTRR

$$\sqrt{k+1}\overline{p}_{k+1}(x|w_{\mathcal{N}[0,1]}) - x.\overline{p}_k(x|w_{\mathcal{N}[0,1]}) + \sqrt{k}\overline{p}_{k-1}(x|w_{\mathcal{N}[0,1]}) = 0, \qquad x \in \mathbb{R},$$

for $k \in \mathbb{N}$, starting from $\overline{p}_0(x|w_{\mathcal{N}[0,1]}) = 1$, $\overline{p}_1(x|w_{\mathcal{N}[0,1]}) = x$. These polynomials are known as Hermite[13] polynomials.

The Hermite polynomials are plotted in Figure 1.1, for orders $k = 2, 5, 8$.

1.4.2.2. Laguerre Polynomials

The standard exponential density function

$$w_{\text{Exp}}(x) = \mathbf{1}(x \geq 0)\exp(-x) \tag{1.18}$$

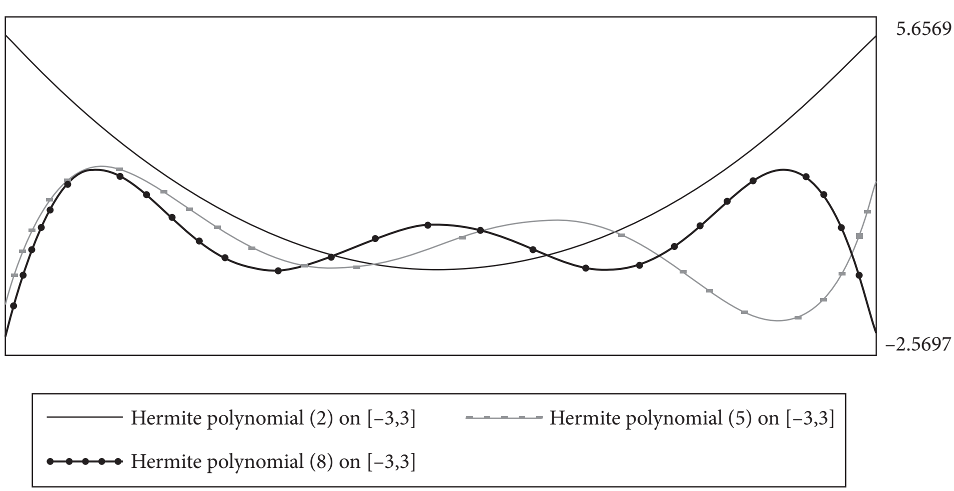


FIGURE 1.1 Hermite polynomials.

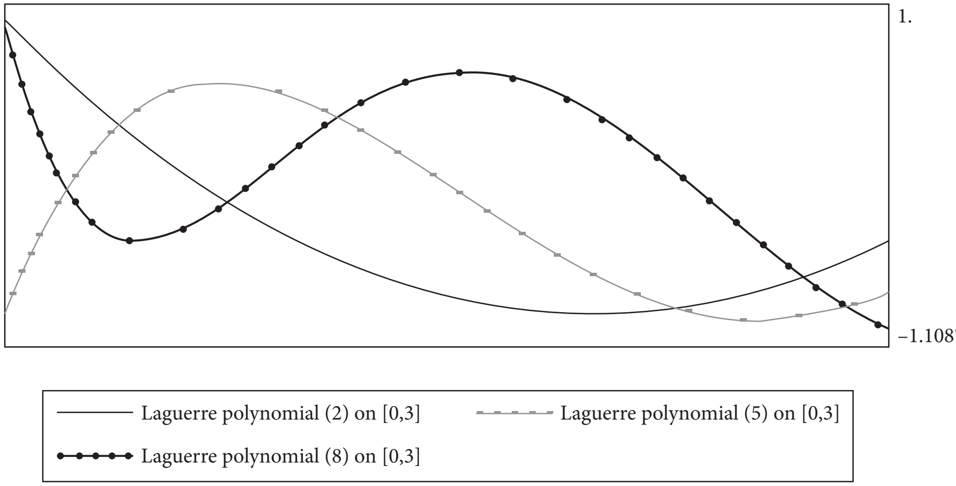


FIGURE 1.2 Laguerre polynomials.

gives rise to the orthonormal Laguerre[14] polynomials, with TTRR

$$(k+1)\,\overline{p}_{k+1}(x|w_{\text{Exp}}) + (2k+1-x)\,\overline{p}_k(x|w_{\text{Exp}}) + k.\overline{p}_{k-1}(x|w_{\text{Exp}}) = 0,\ x \in [0,\infty).$$

for $k \in \mathbb{N}$, starting from $\overline{p}_0(x|w_{\text{Exp}}) = 1$, $\overline{p}_1(x|w_{\text{Exp}}) = x-1$.

These polynomials are plotted in Figure 1.2, for orders $k = 2, 5, 8$.

1.4.2.3. Legendre Polynomials

The uniform density on $[-1,1]$,

$$w_{\mathcal{U}[-1,1]}(x) = \tfrac{1}{2}\mathbf{1}(|x| \leq 1),$$

generates the orthonormal Legendre[15] polynomials on $[-1,1]$, with TTRR

$$\frac{k+1}{\sqrt{2k+3}\sqrt{2k+1}}\overline{p}_{k+1}(x|w_{\mathcal{U}[-1,1]}) - x \cdot \overline{p}_k(x|w_{\mathcal{U}[-1,1]}) + \frac{k}{\sqrt{2k+1}\sqrt{2k-1}}\overline{p}_{k-1}(x|w_{\mathcal{U}[-1,1]}) = 0, \qquad |x| \leq 1,$$

for $k \in \mathbb{N}$, starting from $\overline{p}_0(x|w_{\mathcal{U}[-1,1]}) = 1$, $\overline{p}_1(x|w_{\mathcal{U}[-1,1]}) = \sqrt{3}x$.

Moreover, substituting $x = 2u-1$, it is easy to verify that the uniform density

$$w_{\mathcal{U}[0,1]}(u) = \mathbf{1}(0 \leq u \leq 1)$$

on $[0,1]$ generates the orthonormal polynomials

$$\overline{p}_k(u|w_{\mathcal{U}[0,1]}) = \overline{p}_k(2u-1|w_{\mathcal{U}[-1,1]}),$$

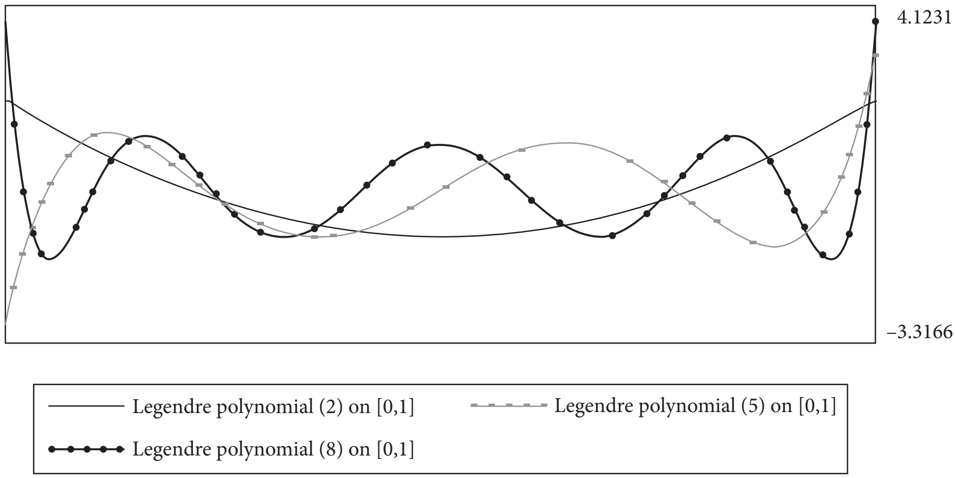


FIGURE 1.3 Shifted Legendre polynomials.

which are known as the shifted Legendre polynomials, also called the Legendre polynomials on the unit interval. The TTRR involved is

$$\frac{(k+1)/2}{\sqrt{2k+3}\sqrt{2k+1}}\overline{p}_{k+1}(u|w_{\mathcal{U}[0,1]}) + (0.5-u)\cdot\overline{p}_k(u|w_{\mathcal{U}[0,1]}) + \frac{k/2}{\sqrt{2k+1}\sqrt{2k-1}}\overline{p}_{k-1}(u|w_{\mathcal{U}[0,1]}) = 0, \qquad 0 \le u \le 1,$$

for $k \in \mathbb{N}$, starting from $\overline{p}_0(u|w_{\mathcal{U}[0,1]}) = 1$, $\overline{p}_1(u|w_{\mathcal{U}[0,1]}) = \sqrt{3}(2u-1)$.

The latter Legendre polynomials are plotted in Figure 1.3, for orders $k = 2, 5, 8$.

1.4.2.4. Chebyshev Polynomials

Chebyshev polynomials are generated by the density function

$$w_{\mathcal{C}[-1,1]}(x) = \frac{1}{\pi\sqrt{1-x^2}}\mathbf{1}(|x|<1), \tag{1.19}$$

with corresponding distribution function

$$W_{\mathcal{C}[-1,1]}(x) = 1 - \pi^{-1}\arccos(x), \qquad x \in [-1,1]. \tag{1.20}$$

The orthogonal (but not orthonormal) Chebyshev polynomials $p_k(x|w_{\mathcal{C}[-1,1]})$ satisfy the TTRR

$$p_{k+1}(x|w_{\mathcal{C}[-1,1]}) - 2xp_k(x|w_{\mathcal{C}[-1,1]}) + p_{k-1}(x|w_{\mathcal{C}[-1,1]}) = 0, \qquad |x|<1, \tag{1.21}$$

for $k \in \mathbb{N}$, starting from $p_0(x|w_{\mathcal{C}[-1,1]}) = 1$, $p_1(x|w_{\mathcal{C}[-1,1]}) = x$, with orthogonality properties

$$\int_{-1}^{1} \frac{p_k(x|w_{\mathcal{C}[-1,1]})p_m(x|w_{\mathcal{C}[-1,1]})}{\pi\sqrt{1-x^2}}\,dx = \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1/2 & \text{if} \quad k = m \in \mathbb{N}, \\ 1 & \text{if} \quad k = m = 0. \end{cases}$$

An important practical difference with the other polynomials discussed so far is that Chebyshev polynomials have the closed form:

$$p_k(x|w_{\mathcal{C}[-1,1]}) = \cos\left(k \cdot \arccos\left(x\right)\right). \tag{1.22}$$

To see this, observe from (1.20) and the well-known sine–cosine formulas that

$$\begin{aligned} &\int_{-1}^{1} \frac{\cos(k \cdot \arccos(x))\cos(m \cdot \arccos(x))}{\pi\sqrt{1-x^2}}\,dx \\ &\quad = -\frac{1}{\pi}\int_{-1}^{1} \cos(k \cdot \arccos(x))\cos(m \cdot \arccos(x))\,d\arccos(x) \\ &\quad = \frac{1}{\pi}\int_{0}^{\pi} \cos(k \cdot \theta)\cos(m \cdot \theta)\,d\theta = \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1/2 & \text{if} \quad k = m \in \mathbb{N}, \\ 1 & \text{if} \quad k = m = 0. \end{cases} \end{aligned}$$

Moreover, it follows from the easy equality $\cos((k+1)\theta) - 2\cos(\theta)\cos(k \cdot \theta) + \cos((k-1)\theta) = 0$ that the functions (1.22) satisfy the TTRR (1.21) and are therefore genuine polynomials, and so are the orthonormal Chebyshev polynomials

$$\overline{p}_k(x|w_{\mathcal{C}[-1,1]}) = \begin{cases} 1 & \text{for} \quad k = 0, \\ \sqrt{2}\cos(k \cdot \arccos(x)) & \text{for} \quad k \in \mathbb{N}. \end{cases}$$

Substituting $x = 2u - 1$ for $u \in [0,1]$ in (1.20) yields

$$W_{\mathcal{C}[0,1]}(u) = 1 - \pi^{-1}\arccos(2u-1) \tag{1.23}$$

with density function

$$w_{\mathcal{C}[0,1]}(u) = \frac{1}{\pi\sqrt{u(1-u)}} \tag{1.24}$$

and shifted orthonormal Chebyshev polynomials

$$\overline{p}_k(u|w_{\mathcal{C}[0,1]}) = \begin{cases} 1 & \text{for} \quad k = 0, \\ \sqrt{2}\cos(k \cdot \arccos(2u-1)) & \text{for} \quad k \in \mathbb{N}. \end{cases} \tag{1.25}$$

The polynomials (1.25) are plotted in Figure 1.4, for orders $k = 2, 5, 8$.

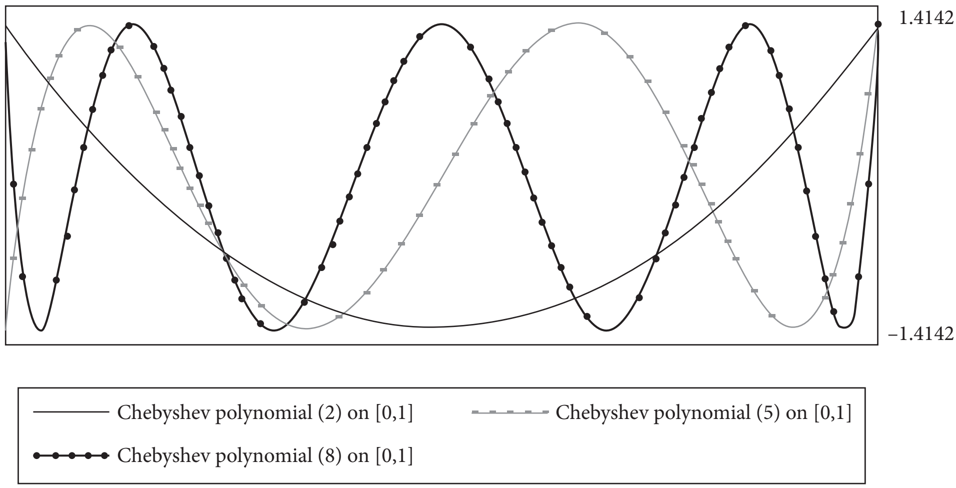


FIGURE 1.4 Shifted Chebyshev polynomials.

1.4.3. Completeness

The reason for considering orthonormal polynomials is the following.

Theorem 1.11. *Let $w(x)$ be a density function on $\mathbb{R}$ satisfying the moment conditions (1.12). Then the orthonormal polynomials $\overline{p}_k(x|w)$ generated by w form a complete orthonormal sequence in the Hilbert space $L^2(w)$. In particular, for any function $f \in L^2(w)$ and with X a random drawing from w,*

$$f(X) = \sum_{k=0}^{\infty} \gamma_k \overline{p}_k(X|w) \qquad \text{a.s.,} \tag{1.26}$$

where $\gamma_k = \int_{-\infty}^{\infty} \overline{p}_m(x|w) f(x) w(x)\, dx$ with $\sum_{k=0}^{\infty} \gamma_k^2 = \int_{-\infty}^{\infty} f(x)^2 w(x)\, dx$.

Proof. Let $f_n(x) = \sum_{m=0}^{n} \gamma_m \overline{p}_m(x|w)$. Then $||f - f_n||^2 = ||f||^2 - \sum_{m=0}^{n} \gamma_m^2$, which is not hard to verify, hence $\sum_{m=0}^{\infty} \gamma_m^2 \le ||f||^2 < \infty$ and thus $\lim_{n\to\infty} \sum_{m=n+1}^{\infty} \gamma_m^2 = 0$. The latter implies that f_n is a Cauchy sequence in $L^2(w)$, with limit $\overline{f} \in \text{span}(\{\overline{p}_m(\cdot|w)\}_{m=0}^{\infty}) \subset L^2(w)$. Thus, $\lim_{n\to\infty} ||\overline{f} - f_n|| = 0$.

To prove the completeness of the sequence $\overline{p}_m(\cdot|w)$, we need to show that $||\overline{f} - f|| = 0$, because then $f \in \text{span}(\{\overline{p}_m(\cdot|w)\}_{m=0}^{\infty})$, which by the arbitrariness of $f \in L^2(w)$ implies that $L^2(w) = \text{span}(\{\overline{p}_m(\cdot|w)\}_{m=0}^{\infty})$. This will be done by showing that for a random drawing X from $w(x)$, we obtain

$$\Pr[\overline{f}(X) = f(X)] = 1, \tag{1.27}$$

because then $||\overline{f} - f||^2 = E[(\overline{f}(X) - f(X))^2] = 0$. In turn, (1.27) is true if for all $t \in \mathbb{R}$, we obtain

$$E[(\overline{f}(X) - f(X)) \exp(i \cdot t \cdot X)] = 0, \tag{1.28}$$

because of the uniqueness of the Fourier transform.[16]

To prove (1.28), note first that the limit function $\overline{f}$ can be written as $\overline{f}(x) = \sum_{m=0}^{n} \gamma_m \overline{p}_m(x|w) + \varepsilon_n(x)$, where $\lim_{n\to\infty} \int_{-\infty}^{\infty} \varepsilon_n(x)^2 w(x)\ dx = 0$. Therefore,

$$\begin{aligned}
&\left| \int_{-\infty}^{\infty} (\overline{f}(x) - f(x)) \overline{p}_m(x|w) w(x)\ dx \right| = \left| \int_{-\infty}^{\infty} \varepsilon_n(x) \overline{p}_m(x|w) w(x)\ dx \right| \\
&\leq \sqrt{\int_{-\infty}^{\infty} \varepsilon_n(x)^2 w(x)\ dx} \sqrt{\int_{-\infty}^{\infty} \overline{p}_m(x|w)^2 w(x)\ dx} = \sqrt{\int_{-\infty}^{\infty} \varepsilon_n(x)^2 w(x)\ dx} \\
&\to 0
\end{aligned}$$

for $n \to \infty$, which implies that for any $g \in \text{span}(\{\overline{p}_m(\cdot|w)\}_{m=0}^{\infty})$, we have $\int_{-\infty}^{\infty} (\overline{f}(x) - f(x)) g(x) w(x)\ dx = 0$. Consequently, $E[(\overline{f}(X) - f(X)) \exp(i \cdot t \cdot X)] = \int_{-\infty}^{\infty} (\overline{f}(x) - f(x)) \exp(i \cdot t \cdot x) w(x)\ dx = 0$ for all $t \in \mathbb{R}$, because it follows from the well-known series expansions of $\cos(t \cdot x) = \text{Re}[\exp(i \cdot t \cdot x)]$ and $\sin(t \cdot x) = \text{Im}[\exp(i \cdot t \cdot x)]$ that these functions are elements of $\text{span}(\{\overline{p}_m(\cdot|w)\}_{m=0}^{\infty})$. Thus, $\{\overline{p}_m(\cdot|w)\}_{m=0}^{\infty}$ is complete in $L^2(w)$. The result (1.26) now follows from Theorem 1.9. ■

1.4.4. Application to the SNP Index Regression Model

Suppose that the response function $f(x)$ in the index regression model (1.1) satisfies

$$\sup_x |f(x)| \cdot \exp(-t_0 \cdot |x|) = M(t_0) < \infty \qquad \text{for some } t_0 > 0. \tag{1.29}$$

so that $-M(t_0) \exp(t_0 \cdot |x|) \leq f(x) \leq M(t_0) \exp(t_0 \cdot |x|)$. Then for the standard normal density $w_{\mathcal{N}[0,1]}(x)$, we have $\int_{-\infty}^{\infty} f(x)^2 w_{\mathcal{N}[0,1]}(x)\ dx < 2M(t_0) \exp(t_0^2/2) < \infty$; hence $f \in L^2(w_{\mathcal{N}[0,1]})$, so that $f(x)$ has the Hermite series expansion

$$f(x) = \sum_{m=0}^{\infty} \delta_{0,m} \overline{p}_m(x|w_{\mathcal{N}[0,1]}) = \delta_{0,0} + \delta_{0,1} x + \sum_{k=2}^{\infty} \delta_{0,k} \overline{p}_k(x|w_{\mathcal{N}[0,1]}) \qquad \text{a.e. on } \mathbb{R},$$

with $\delta_{0,m} = \int_{-\infty}^{\infty} f(x) \overline{p}_m(x|w_{\mathcal{N}[0,1]}) w_{\mathcal{N}[0,1]}(x)\ dx$ for $m = 0, 1, 2, \ldots$. Thus, model (1.1) now reads

$$E[Y|X] = \lim_{n\to\infty} f_n(X_1 + \theta_0' X_2 | \boldsymbol{\delta}_n^0) \qquad \text{a.s,} \tag{1.30}$$

where

$$f_n(x|\boldsymbol{\delta}_n) = \delta_0 + \delta_1 x + \sum_{k=2}^{n} \delta_k \overline{p}_k(x|w_{\mathcal{N}[0,1]}) \tag{1.31}$$

with $\boldsymbol{\delta}_n = (\delta_0, \delta_1, \ldots, \delta_n, 0, 0, 0, \ldots)$ and $\boldsymbol{\delta}_n^0 = (\delta_{0,0}, \delta_{0,1}, \ldots, \delta_{0,n}, 0, 0, 0, \ldots)$.

For fixed $n \in \mathbb{N}$ the parameters involved can be approximated by weighted nonlinear regression of Y on $f_n(X_1 + \theta' X_2|\boldsymbol{\delta}_n)$, given a random sample $\{(Y_j, X_j)\}_{j=1}^N$ from (Y, X) and given predefined compact parameter spaces Δ_n and Θ for $\boldsymbol{\delta}_n^0$ and θ_0, respectively. Then the weighted NLLS sieve estimator of $(\theta_0, \boldsymbol{\delta}_n^0)$ is

$$\left(\widehat{\theta}_n, \widehat{\boldsymbol{\delta}}_n\right) = \arg \min_{(\theta, \boldsymbol{\delta}_n) \in \Theta \times \Delta_n} \frac{1}{N} \sum_{j=1}^N \left(Y_j - f_n(X_{1,j} + \theta' X_{2,j}|\boldsymbol{\delta}_n)\right)^2 K(||X_j||), \tag{1.32}$$

where $K(x)$ is a positive weight function on $(0, \infty)$ satisfying $\sup_{x>0} x^n K(x) < \infty$ for all $n \geq 0$. The reason for this weight function is to guarantee that

$$E\left[\sup_{(\theta, \boldsymbol{\delta}_n) \in \Theta \times \Delta_n} \left(Y - f_n(X_1 + \theta' X_2|\boldsymbol{\delta}_n)\right)^2 K(||X||)\right] < \infty$$

without requiring that $E[||X||^{2n}] < \infty$. Then by Jennrich's (1969) uniform law of large numbers and for fixed n, we have

$$\sup_{(\theta, \boldsymbol{\delta}_n) \in \Theta \times \Delta_n} \left| \frac{1}{N} \sum_{j=1}^N \left(Y_j - f_n(X_{1,j} + \theta' X_{2,j}|\boldsymbol{\delta}_n)\right)^2 K(||X_j||) - g_n(\theta, \boldsymbol{\delta}_n) \right|$$
$$= o_p(1),$$

where $g_n(\theta, \boldsymbol{\delta}_n) = E[(Y - f_n(X_1 + \theta' X_2|\boldsymbol{\delta}_n))^2 K(||X||)]$, so that

$$p \lim_{N \to \infty} \left(\widehat{\theta}_n, \widehat{\boldsymbol{\delta}}_n\right) = \left(\overline{\theta}_n, \overline{\boldsymbol{\delta}}_n\right) = \arg \min_{(\theta, \boldsymbol{\delta}_n) \in \Theta \times \Delta_n} g_n(\theta, \boldsymbol{\delta}_n).$$

Under some alternative conditions the same result can be obtained by using the Wald (1949) consistency result in van der Vaart (1998, Theorem 5.14), which does not require that the expectation of the objective function is finite for all values of the parameters, so that in that case there is no need for the weight function $K(x)$.

Note that, in general, $\overline{\theta}_n \neq \theta_0$. Nevertheless, it can be shown that under some additional regularity conditions,[17] and with $n = n_N$ an arbitrary subsequence of N, $p\lim_{N \to \infty} \widehat{\theta}_{n_N} = \theta_0$ and $p\lim_{N \to \infty} \int_{-\infty}^{\infty} (f_{n_N}(x|\widehat{\boldsymbol{\delta}}_{n_N}) - f(x))^2 w_{\mathcal{N}[0,1]}(x)\, dx = 0$.

1.5. Non-Polynomial Complete Orthonormal Sequences

Recall that the support of a density $w(x)$ on $\mathbb{R}$ is defined as the set $\{x \in \mathbb{R} : w(x) > 0\}$. For example, the support of the standard exponential density (1.18) is the interval $[0, \infty)$. In this chapter I will only consider densities $w(x)$ with connected support—that is, the support is an interval—and for notational convenience this support will be denoted by an *open* interval (a, b), where $a = \inf_{w(x)>0} x \geq -\infty$ and $b = \sup_{w(x)>0} x \leq \infty$, even if for finite a and/or b, $w(a) > 0$ or $w(b) > 0$.

1.5.1. Nonpolynomial Sequences Derived from Polynomials

For every density $w(x)$ with support (a,b), $\int_a^b f(x)^2\,dx<\infty$ implies that $f(x)/\sqrt{w(x)}\in L^2(w)$. Therefore, the following corollary of Theorem 1.11 holds trivially.

Theorem 1.12. *Every function $f\in L^2(a,b)$ can be written as*

$$f(x)=\sqrt{w(x)}\left(\sum_{k=0}^{\infty}\gamma_k\overline{p}_k(x|w)\right)\qquad a.e.\ on\ (a,b),$$

where w is a density with support (a,b) satisfying the moment conditions (1.12) and $\gamma_k=\int_a^b f(x)\overline{p}_k(x|w)\sqrt{w(x)}\,dx$. Consequently, $L^2(a,b)$ is a Hilbert space with complete orthonormal sequence $\psi_k(x|w)=\overline{p}_k(x|w)\sqrt{w(x)}$, $k\in\mathbb{N}_0$.

If (a,b) is bounded, then there is another way to construct a complete orthonormal sequence in $L^2(a,b)$, as follows. Let $W(x)$ be the distribution function of a density w with bounded support (a,b). Then $G(x)=a+(b-a)W(x)$ is a one-to-one mapping of (a,b) onto (a,b), with inverse $G^{-1}(y)=W^{-1}((y-a)/(b-a))$, where W^{-1} is the inverse of $W(x)$. For every $f\in L^2(a,b)$, we have

$$(b-a)\int_a^b f(G(x))^2\,w(x)\,dx=\int_a^b f(G(x))^2\,dG(x)=\int_a^b f(x)^2\,dx<\infty.$$

Hence $f(G(x))\in L^2(w)$ and thus by Theorem 1.11 we have $f(G(x))=\sum_{k=0}^{\infty}\gamma_k\overline{p}_k(x|w)$ a.e. on (a,b), where $\gamma_k=\int_a^b f(G(x))\overline{p}_k(x|w)w(x)\,dx$. Consequently,

$$f(x)=f\left(G\left(G^{-1}(x)\right)\right)=\sum_{k=0}^{\infty}\gamma_k\overline{p}_k\left(G^{-1}(x)\,|w\right)\qquad \text{a.e. on } (a,b).$$

Note that $dG^{-1}(x)/dx=dG^{-1}(x)/dG(G^{-1}(x))=1/G'(G^{-1}(x))$, so that

$$\begin{aligned}
&\int_a^b \overline{p}_k\left(G^{-1}(x)\,|w\right)\overline{p}_m\left(G^{-1}(x)\,|w\right)dx\\
&=\int_a^b \overline{p}_k\left(G^{-1}(x)\,|w\right)\overline{p}_m\left(G^{-1}(x)\,|w\right)G'\left(G^{-1}(x)\right)dG^{-1}(x)\\
&=\int_a^b \overline{p}_k(x|w)\,\overline{p}_m(x|w)\,G'(x)\,dx\\
&=(b-a)\int_a^b \overline{p}_k(x|w)\,\overline{p}_m(x|w)\,w(x)\,dx=(b-a)\,\mathbf{1}(k=m).
\end{aligned}$$

Thus, we have the following theorem.

Theorem 1.13. *Let w be a density with bounded support (a,b) and let W be the c.d.f. of w, with inverse W^{-1}. Then the functions*

$$\psi_k(x|w) = \overline{p}_k\left(W^{-1}((x-a)/(b-a))|w\right)/\sqrt{(b-a)}, \qquad k \in \mathbb{N}_0$$

form a complete orthonormal sequence in $L^2(a,b)$. Hence, every function $f \in L^2(a,b)$ has the series representation $f(x) = \sum_{k=0}^{\infty} \gamma_k \psi_k(x|w)$ a.e. on (a,b), with $\gamma_k = \int_a^b \psi_k(x|w) f(x)\,dx$.

1.5.2. Trigonometric Sequences

Let us specialize the result in Theorem 1.13 to the case of the Chebyshev polynomials on $[0,1]$, with $a=0$, $b=1$ and W, w and $\overline{p}_k(u|w)$ given by (1.23), (1.24), and (1.25), respectively. Observe that in this case $W_{C[0,1]}^{-1}(u) = (1-\cos(\pi u))/2$. It follows now straightforwardly from (1.25) and the easy equality $\arccos(-x) = \pi - \arccos(x)$ that for $k \in \mathbb{N}$, $\overline{p}_k(W_{C[0,1]}^{-1}(u)|w_{C[0,1]}) = \sqrt{2}\cos(k\pi)\cos(k\pi u) = \sqrt{2}(-1)^k \cos(k\pi u)$, which by Theorem 1.13 implies the following.

Theorem 1.14. *The cosine sequence*

$$\psi_k(u) = \begin{cases} 1 & \text{for} \quad k=0, \\ \sqrt{2}\cos(k\pi u) & \text{for} \quad k \in \mathbb{N}, \end{cases}$$

is a complete orthonormal sequence in $L^2(0,1)$. Hence, every function $f \in L^2(0,1)$ has the series representation $f(u) = \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2}\cos(k\pi u)$ a.e. on $(0,1)$, with $\gamma_0 = \int_0^1 f(u)\,du$, $\gamma_k = \sqrt{2}\int_0^1 \cos(k\pi u) f(u)\,du$ for $k \in \mathbb{N}$.

This result is related to classical Fourier analysis. Consider the following sequence of functions on $[-1,1]$:

$$\varphi_0(x) = 1,$$
$$\varphi_{2k-1}(x) = \sqrt{2}\sin(k\pi x),\ \varphi_{2k}(x) = \sqrt{2}\cos(k\pi x), \qquad k \in \mathbb{N}. \tag{1.33}$$

These functions are know as the Fourier series on $[-1,1]$. It is easy to verify that these functions are orthonormal with respect to the uniform density $w_{\mathcal{U}[-1,1]}(x) = \frac{1}{2}\mathbf{1}(|x| \leq 1)$ on $[-1,1]$, that is, $\frac{1}{2}\int_{-1}^{1} \varphi_m(x)\varphi_k(x)\,dx = \mathbf{1}(m=k)$. The following theorem is a classical Fourier analysis result.

Theorem 1.15. *The Fourier sequence $\{\varphi_n\}_{n=0}^{\infty}$ is complete in $L^2(w_{\mathcal{U}[-1,1]})$.* [18]

Now Theorem 1.14 is a corollary of Theorem 1.15. To see this, let $f \in L^2(0,1)$ be arbitrary. Then $g(x) = f(|x|) \in L^2(w_{\mathcal{U}[-1,1]})$; hence

$$g(x) = \alpha + \sum_{k=1}^{\infty} \beta_k \sqrt{2}\cos(k\pi x) + \sum_{m=1}^{\infty} \gamma_m \sqrt{2}\sin(k\pi x)$$

a.e. on $[-1,1]$, where

$$\alpha = \frac{1}{2}\int_{-1}^{1} g(x)\, dx = \int_{0}^{1} f(u)\, du$$

$$\beta_k = \frac{1}{2}\int_{-1}^{1} g(x)\sqrt{2}\cos(k\pi x)\ dx = \int_{0}^{1} f(u)\sqrt{2}\cos(k\pi u)\ du$$

$$\gamma_m = \frac{1}{2}\int_{-1}^{1} g(x)\sqrt{2}\sin(k\pi x)\ dx = 0$$

so that $f(u) = \alpha + \sum_{k=1}^{\infty}\beta_k\sqrt{2}\cos(k\pi u)$ a.e. on $[0,1]$.

Similarly, given an arbitrary $f \in L^2(0,1)$, let $g(x) = (\mathbf{1}(x \geq 0) - \mathbf{1}(x < 0))f(|x|)$. Then $g(x) = \sum_{m=1}^{\infty}\gamma_m\sqrt{2}\sin(k\pi x)$ a.e. on $[-1,1]$; hence $f(u) = \sum_{m=1}^{\infty}\gamma_m\sqrt{2}\sin(k\pi u)$ a.e. on $(0,1)$, where $\gamma_m = \int_0^1 f(u)\sqrt{2}\sin(m\pi u)\ du$. Therefore, we have the following corollary.

Corollary 1.1. *The sine sequence* $\sqrt{2}\sin(m\pi u)$, $m \in \mathbb{N}$, *is complete in* $L^2(0,1)$.

Although this result implies that for every $f \in L^2(0,1)$, $\lim_{n\to\infty} f_n(u) = f(u)$ a.e. on $(0,1)$, where $f_n(u) = \sum_{m=1}^{n}\gamma_m\sqrt{2}\sin(k\pi u)$ with $\gamma_m = \sqrt{2}\int_0^1 f(u)\sin(m\pi u)\ du$, the approximation $f_n(u)$ may be very poor in the tails of $f(u)$ if $f(0) \neq 0$ and $f(1) \neq 0$, because, in general, $\lim_{u\downarrow 0}\lim_{n\to\infty} f_n(u) \neq \lim_{n\to\infty}\lim_{u\downarrow 0} f_n(u)$, and similarly for $u \uparrow 1$. Therefore, the result of Corollary 1.1 is of limited practical significance.

1.6. Density and Distribution Functions

1.6.1. General Univariate SNP Density Functions

Let $w(x)$ be a density function with support (a,b). Then for any density $f(x)$ on (a,b), we obtain

$$g(x) = \sqrt{f(x)}/\sqrt{w(x)} \in L^2(w), \tag{1.34}$$

with $\int_a^b g(x)^2 w(x)\ dx = \int_a^b f(x)\ dx = 1$. Therefore, given a complete orthonormal sequence $\{\rho_m\}_{m=0}^{\infty}$ in $L^2(w)$ with $\rho_0(x) \equiv 1$ and denoting $\gamma_m = \int_a^b \rho_m(x)g(x)w(x)\ dx$, any density $f(x)$ on (a,b) can be written as

$$f(x) = w(x)\left(\sum_{m=0}^{\infty}\gamma_m\rho_m(x)\right)^2 \quad \text{a.e. on } (a,b), \qquad \text{with } \sum_{m=0}^{\infty}\gamma_m^2 = \int_a^b f(x)\ dx = 1. \tag{1.35}$$

The reason for the square in (1.35) is to guarantee that $f(x)$ is non-negative.

A problem with the series representation (1.35) is that in general the parameters involved are not unique. To see this, note that if we replace the function $g(x)$ in (1.34) by $g_B(x) = (\mathbf{1}(x \in B) - \mathbf{1}(x \in (a,b)\backslash B))\sqrt{f(x)}/\sqrt{w(x)}$, where B is an arbitrary Borel set, then $g_B(x) \in L^2(w)$ and $\int_a^b g_B(x)^2 w(x)\, dx = \int_a^b f(x)\, dx = 1$, so that (1.35) also holds for the sequence

$$\gamma_m = \int_a^b \rho_m(x) g_B(x) w(x)\, dx$$
$$= \int_{(a,b)\cap B} \rho_m(x)\sqrt{f(x)}\sqrt{w(x)}\, dx - \int_{(a,b)\backslash B} \rho_m(x)\sqrt{f(x)}\sqrt{w(x)}\, dx.$$

In particular, using the fact that $\rho_0(x) \equiv 1$, we obtain

$$\gamma_0 = \int_{(a,b)\cap B} \sqrt{f(x)}\sqrt{w(x)}\, dx - \int_{(a,b)\backslash B} \sqrt{f(x)}\sqrt{w(x)}\, dx,$$

so that the sequence γ_m in (1.35) is unique if γ_0 is maximal. In any case we may without loss of generality assume that $\gamma_0 \in (0,1)$, so that without loss of generality the γ_m's can be reparameterized as

$$\gamma_0 = \frac{1}{\sqrt{1 + \sum_{k=1}^{\infty} \delta_k^2}}, \qquad \gamma_m = \frac{\delta_m}{\sqrt{1 + \sum_{k=1}^{\infty} \delta_k^2}},$$

where $\sum_{k=1}^{\infty} \delta_k^2 < \infty$. This reparameterization does not solve the lack of uniqueness problem, of course, but is convenient in enforcing the restriction $\sum_{m=0}^{\infty} \gamma_m^2 = 1$.

On the other hand, under certain conditions on $f(x)$ the δ_m's are unique, as will be shown in Section 1.6.4.

Summarizing, the following result has been shown.

Theorem 1.16. *Let $w(x)$ be a univariate density function with support (a,b), and let $\{\rho_m\}_{m=0}^{\infty}$ be a complete orthonormal sequence in $L^2(w)$, with $\rho_0(x) \equiv 1$. Then for any density $f(x)$ on (a,b) there exist possibly uncountably many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty} \delta_m^2 < \infty$ such that*

$$f(x) = \frac{w(x)\left(1 + \sum_{m=1}^{\infty} \delta_m \rho_m(x)\right)^2}{1 + \sum_{m=1}^{\infty} \delta_m^2} \qquad \text{a.e. on } (a,b). \tag{1.36}$$

Moreover, for the sequence $\{\delta_m\}_{m=1}^{\infty}$ for which $\sum_{m=1}^{\infty} \delta_m^2$ is minimal,

$$\sqrt{f(x)} = \frac{\sqrt{w(x)}\left(1 + \sum_{m=1}^{\infty} \delta_m \rho_m(x)\right)}{\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2}} \qquad \text{a.e. on } (a,b);$$

hence

$$\delta_m = \frac{\int_a^b \rho_m(x)\sqrt{f(x)}\sqrt{w(x)}\, dx}{\int_a^b \sqrt{f(x)}\sqrt{w(x)}\, dx}, \qquad m \in \mathbb{N}. \tag{1.37}$$

In practice, the result of Theorem 1.16 cannot be used directly in SNP modeling, because it is impossible to estimate infinitely many parameters. Therefore, the density (1.36) is usually approximated by

$$f_n(x) = \frac{w(x)\left(1+\sum_{m=1}^{n}\delta_m\rho_m(x)\right)^2}{1+\sum_{m=1}^{n}\delta_m^2} \tag{1.38}$$

for some natural number n, possibly converging to infinity with the sample size. Following Gallant and Nychka (1987), I will refer to truncated densities of the type (1.38) as *SNP densities.*

Obviously,

Corollary 1.2. *Under the conditions of Theorem 1.16,* $\lim_{n\to\infty} f_n(x) = f(x)$ *a.e. on* (a,b). *Moreover, it is not hard to verify that*

$$\int_a^b |f(x)-f_n(x)|\,dx \le 4\sqrt{\sum_{m=n+1}^{\infty}\delta_m^2} + 2\sum_{m=n+1}^{\infty}\delta_m^2 = o(1), \tag{1.39}$$

where the δ_m*'s are given by* (1.37), *so that with* $F(x)$ *the c.d.f. of* $f(x)$ *and* $F_n(x)$ *the c.d.f. of* $f_n(x)$, *we obtain*

$$\lim_{n\to\infty}\sup_x |F(x)-F_n(x)| = 0.$$

Remarks

1. The rate of convergence to zero of the tail sum $\sum_{m=n+1}^{\infty}\delta_m^2$ depends on the smoothness, or the lack thereof, of the density $f(x)$. Therefore, the question of how to choose the truncation order n given an *a priori* chosen approximation error cannot be answered in general.
2. In the case that the $\rho_m(x)$'s are polynomials, the SNP density $f_n(x)$ has to be computed recursively via the corresponding TTRR (1.17), except in the case of Chebyshev polynomials, but that is not much of a computational burden. However, the computation of the corresponding SNP distribution function $F_n(x)$ is more complicated. See, for example, Stewart (2004) for SNP distribution functions on $\mathbb{R}$ based on Hermite polynomials, and see Bierens (2008) for SNP distribution functions on $[0,1]$ based on Legendre polynomials. Both cases require to recover the coefficients $\ell_{m,k}$ of the polynomials $\overline{p}_k(x|w) = \sum_{m=0}^{k}\ell_{m,k}x^m$, which can be done using the TTRR involved. Then with $P_n(x|w) = (1, \overline{p}_1(x|w), \ldots, \overline{p}_n(x|w))'$, $Q_n(x) = (1, x, \ldots, x^n)'$, $\delta = (1, \delta_1, \ldots, \delta_n)$, and L_n the lower-triangular matrix consisting of the coefficients $\ell_{m,k}$, we can write $f_n(x) = (\delta'\delta)^{-1}w(x)(\delta' P_n(x|w))^2 = (\delta'\delta)^{-1}\delta' L_n Q_n(x) Q_n(x)' w(x) L_n'\delta$; hence

$$F_n(x) = \frac{1}{\delta'\delta}\delta' L_n\left(\int_{-\infty}^{x} Q_n(z)Q_n(z)'w(z)\,dz\right)L_n'\delta = \frac{\delta' L_n M_n(x) L_n'\delta}{\delta'\delta},$$

where $M_n(x)$ is the $(n+1)\times(n+1)$ matrix with typical elements $\int_{-\infty}^{x} z^{i+j}w(z)\,dz$ for $i,j = 0,1,\ldots,n$. This is the approach proposed by Bierens (2008). The approach in Stewart (2004) is in essence the same and is therefore equally cumbersome.

1.6.2. Bivariate SNP Density Functions

Now let $w_1(x)$ and $w_2(y)$ be a pair of density functions on $\mathbb{R}$ with supports (a_1,b_1) and (a_2,b_2), respectively, and let $\{\rho_{1,m}\}_{m=0}^{\infty}$ and $\{\rho_{2,m}\}_{m=0}^{\infty}$ be complete orthonormal sequences in $L^2(w_1)$ and $L^2(w_2)$, respectively. Moreover, let $g(x,y)$ be a Borel measurable real function on $(a_1,b_1)\times(a_2,b_2)$ satisfying

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} g(x,y)^2 w_1(x)w_2(y)\,dx\,dy < \infty. \tag{1.40}$$

The latter implies that $g_2(y) = \int_{a_1}^{b_1} g(x,y)^2 w_1(x)\;dx < \infty$ a.e. on (a_2,b_2), so that for each $y\in(a_2,b_2)$ for which $g_2(y)<\infty$ we have $g(x,y)\in L^2(w_1)$. Then $g(x,y) = \sum_{m=0}^{\infty}\gamma_m(y)\rho_{1,m}(x)$ a.e. on (a_1,b_1), where $\gamma_m(y) = \int_{a_1}^{b_1} g(x,y)\rho_{1,m}(x)w_1(x)\,dx$ with $\sum_{m=0}^{\infty}\gamma_m(y)^2 = \int_{a_1}^{b_1} g(x,y)^2\cdot w_1(x)\;dx = g_2(y)$. Because by (1.40) we have $\int_{a_2}^{b_2} g_2(y)w_2(y)\;dy<\infty$, it follows now that for each $y\in(a_2,b_2)$ for which $g_2(y)<\infty$ and all integers $m\geq 0$ we have $\gamma_m(y)\in L^2(w_2)$, so that $\gamma_m(y) = \sum_{k=0}^{\infty}\gamma_{m,k}\rho_{2,k}(y)$ a.e. on (a_2,b_2), where $\gamma_{m,k} = \int_{a_1}^{b_1}\int_{a_2}^{b_2} g(x,y)\rho_{1,m}(x)\rho_{2,k}(y)w_1(x)w_2(y)\,dxdy$ with $\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\gamma_{m,k}^2<\infty$. Hence,

$$g(x,y) = \sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\gamma_{m,k}\rho_{1,m}(x)\rho_{2,k}(y) \qquad \text{a.e. on } (a_1,b_1)\times(a_2,b_2). \tag{1.41}$$

Therefore, it follows similar to Theorem 1.16 that the next theorem holds.

Theorem 1.17. *Given a pair of density functions $w_1(x)$ and $w_2(y)$ with supports (a_1,b_1) and (a_2,b_2), respectively, and given complete orthonormal sequences $\{\rho_{1,m}\}_{m=0}^{\infty}$ and $\{\rho_{2,m}\}_{m=0}^{\infty}$ in $L^2(w_1)$ and $L^2(w_2)$, respectively, with $\rho_{1,0}(x)=\rho_{2,0}(y)\equiv 1$, for every bivariate density $f(x,y)$ on $(a_1,b_1)\times(a_2,b_2)$ there exist possibly uncountably many double arrays $\delta_{m,k}$ satisfying $\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\delta_{m,k}^2<\infty$, with $\delta_{0,0}=1$ by normalization, such that a.e. on $(a_1,b_1)\times(a_2,b_2)$, we obtain*

$$f(x,y) = \frac{w_1(x)w_2(y)\left(\sum_{m=0}^{\infty}\sum_{k=0}^{\infty}\delta_{m,k}\rho_{1,m}(x)\rho_{2,k}(y)\right)^2}{\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\delta_{k,m}^2}.$$

For example, let $w_1(x)$ and $w_2(y)$ be standard normal densities and $\rho_{1,m}(x)$ and $\rho_{2,k}(y)$ Hermite polynomials, that is, $\rho_{1,k}(x)=\rho_{2,k}(x)=\overline{p}_k(x|w_{\mathcal{N}[0,1]})$. Then for any density function $f(x,y)$ on $\mathbb{R}^2$ there exists a double array $\delta_{m,k}$ and associated sequence of SNP densities

$$f_n(x,y) = \frac{\exp\left(-(x^2+y^2)/2\right)}{2\pi\sum_{k=0}^{n}\sum_{m=0}^{n}\delta_{k,m}^2}\left(\sum_{m=0}^{n}\sum_{k=0}^{n}\delta_{m,k}\overline{p}_m(x|w_{\mathcal{N}[0,1]})\overline{p}_k(y|w_{\mathcal{N}[0,1]})\right)^2$$

such that $\lim_{n\to\infty} f_n(x,y) = f(x,y)$ a.e. on $\mathbb{R}^2$.

This result is used by Gallant and Nychka (1987) to approximate the bivariate error density of the latent variable equations in Heckman's (1979) sample selection model.

1.6.3. SNP Densities and Distribution Functions on $[0,1]$

Since the seminal paper by Gallant and Nychka (1987), SNP modeling of density and distribution functions on $\mathbb{R}$ via the Hermite expansion has become the standard approach in econometrics, despite the computational burden of computing the SNP distribution function involved.

However, there is an easy trick to avoid this computational burden, by mapping one-to-one any absolutely continuous distribution function $F(x)$ on (a,b) with density $f(x)$ to an absolutely continuous distribution function $H(u)$ with density $h(u)$ on the unit interval, as follows. Let $G(x)$ be an *a priori* chosen absolutely continuous distribution function with density $g(x)$ and support (a,b). Then we can write

$$F(x) = H(G(x)) \quad \text{and} \quad f(x) = h(G(x))\cdot g(x), \tag{1.42}$$

where

$$H(u) = F(G^{-1}(u)) \quad \text{and} \quad h(u) = f(G^{-1}(u))/g(G^{-1}(u)) \tag{1.43}$$

with $G^{-1}(u)$ the inverse of $G(x)$.

For example, let $(a,b)=\mathbb{R}$ and choose for $G(x)$ the logistic distribution function $G(x) = 1/(1+\exp(-x))$. Then $g(x) = G(x)(1-G(x))$ and $G^{-1}(u) = \ln(u/(1-u))$, hence $h(u) = f(\ln(u/(1-u)))/(u(1-u))$. Similarly, if $(a,b) = (0,\infty)$ and $G(x) = 1-\exp(-x)$, then any density $f(x)$ on $(0,\infty)$ corresponds uniquely to $h(u) = f(\ln(1/(1-u)))/(1-u)$.

The reason for this transformation is that there exist closed-form expressions for SNP densities on the unit interval and their distribution functions. In particular, Theorem 1.18 follows from (1.23)–(1.25) and Corollary 1.2.

Theorem 1.18. *For every density $h(u)$ on $[0,1]$ with corresponding c.d.f. $H(u)$ there exist possibly uncountably many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty}\delta_m^2<\infty$ such that $h(u) = \lim_{n\to\infty} h_n(u)$ a.e. on $[0,1]$, where*

$$h_n(u) = \frac{1}{\pi\sqrt{u(1-u)}}\frac{\left(1+\sum_{m=1}^{n}(-1)^m\delta_m\sqrt{2}\cos(m\cdot\arccos(2u-1))\right)^2}{1+\sum_{m=1}^{n}\delta_m^2}, \tag{1.44}$$

and $\lim_{n\to\infty}\sup_{0\le u\le 1}|\underline{H}_n(1-\pi^{-1}\arccos(2u-1))-H(u)|=0$, *where*

$$\begin{aligned}\underline{H}_n(u) = u + \frac{1}{1+\sum_{m=1}^n \delta_m^2}\Bigg[& 2\sqrt{2}\sum_{k=1}^n \delta_k \frac{\sin(k\pi u)}{k\pi} + \sum_{m=1}^n \delta_m^2 \frac{\sin(2m\pi u)}{2m\pi} \\ & + 2\sum_{k=2}^n\sum_{m=1}^{k-1}\delta_k\delta_m \frac{\sin((k+m)\pi u)}{(k+m)\pi} \\ & + 2\sum_{k=2}^n\sum_{m=1}^{k-1}\delta_k\delta_m \frac{\sin((k-m)\pi u)}{(k-m)\pi}\Bigg]. \end{aligned} \tag{1.45}$$

Moreover, with $w(x)$ being the uniform density on $[0,1]$ and $\rho_m(x)$ being the cosine sequence, it follows from Corollary 1.2 that the next theorem holds.

Theorem 1.19. *For every density $h(u)$ on $[0,1]$ with corresponding c.d.f. $H(u)$ there exist possibly uncountably many sequences $\{\delta_m\}_{m=1}^\infty$ satisfying $\sum_{m=1}^\infty \delta_m^2 < \infty$ such that a.e. on $[0,1]$, $h(u)=\lim_{n\to\infty} h_n(u)$, where*

$$h_n(u) = \frac{\left(1+\sum_{m=1}^n \delta_m \sqrt{2}\cos(m\pi u)\right)^2}{1+\sum_{m=1}^n \delta_m^2}, \tag{1.46}$$

and $\lim_{n\to\infty}\sup_{0\le u\le 1}|\underline{H}_n(u)-H(u)|=0$, *where $\underline{H}_n(u)$ is defined by* (1.45).

The latter follows straightforwardly from (1.46) and the well-known equality $\cos(a)\cos(b) = (\cos(a+b)+\cos(a-b))/2$, and the same applies to the result for $H(u)$ in Theorem 1.18.

1.6.4. Uniqueness of the Series Representation

The density $h(u)$ in Theorem 1.19 can be written as $h(u)=\eta(u)^2/\int_0^1 \eta(v)^2\,dv$, where

$$\eta(u) = 1 + \sum_{m=1}^\infty \delta_m \sqrt{2}\cos(m\pi u) \qquad \text{a.e. on } (0,1). \tag{1.47}$$

Moreover, recall that in general we have

$$\delta_m = \frac{\int_0^1 (\mathbf{1}(u\in B) - \mathbf{1}(u\in[0,1]\backslash B))\sqrt{2}\cos(m\pi u)\sqrt{h(u)}\,du}{\int_0^1 (\mathbf{1}(u\in B) - \mathbf{1}(u\in[0,1]\backslash B))\sqrt{h(u)}\,du},$$

$$\frac{1}{\sqrt{1+\sum_{m=1}^\infty \delta_m^2}} = \int_0^1 (\mathbf{1}(u\in B) - \mathbf{1}(u\in[0,1]\backslash B))\sqrt{h(u)}\,du.$$

for some Borel set B satisfying $\int_0^1 (\mathbf{1}(u \in B) - \mathbf{1}(u \in [0,1]\backslash B))\sqrt{h(u)}\, du > 0$; hence

$$\eta(u) = (\mathbf{1}(u \in B) - \mathbf{1}(u \in [0,1]\backslash B))\sqrt{h(u)}\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2} \tag{1.48}$$

Similarly, given this Borel set B and the corresponding δ_m's, the SNP density (1.46) can be written as $h_n(u) = \eta_n(u)^2 / \int_0^1 \eta_n(v)^2 dv$, where

$$\begin{aligned} \eta_n(u) &= 1 + \sum_{m=1}^{n} \delta_m \sqrt{2} \cos(m\pi u) \\ &= (\mathbf{1}(u \in B) - \mathbf{1}(u \in [0,1]\backslash B))\sqrt{h_n(u)}\sqrt{1 + \sum_{m=1}^{n} \delta_m^2}. \end{aligned} \tag{1.49}$$

Now suppose that $h(u)$ is continuous and positive on $(0,1)$. Moreover, let $S \subset [0,1]$ be the set with Lebesgue measure zero on which $h(u) = \lim_{n\to\infty} h_n(u)$ fails to hold. Then for any $u_0 \in (0,1)\backslash S$ we have $\lim_{n\to\infty} h_n(u_0) = h(u_0) > 0$; hence for sufficient large n we have $h_n(u_0) > 0$. Because obviously $h_n(u)$ and $\eta_n(u)$ are continuous on $(0,1)$, for such an n there exists a small $\varepsilon_n(u_0) > 0$ such that $h_n(u) > 0$ for all $u \in (u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0,1)$, and therefore

$$\mathbf{1}(u \in B) - \mathbf{1}(u \in [0,1]\backslash B) = \frac{\eta_n(u)}{\sqrt{h_n(u)}\sqrt{1 + \sum_{m=1}^{n} \delta_m^2}} \tag{1.50}$$

is continuous on $(u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0,1)$. Substituting (1.50) in (1.48), it follows now that $\eta(u)$ is continuous on $(u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0,1)$; hence by the arbitrariness of $u_0 \in (0,1)/S$, $\eta(u)$ is continuous on $(0,1)$.

Next, suppose that $\eta(u)$ takes positive and negative values on $(0,1)$. Then by the continuity of $\eta(u)$ on $(0,1)$ there exists a $u_0 \in (0,1)$ for which $\eta(u_0) = 0$ and thus $h(u_0) = 0$, which, however, is excluded by the condition that $h(u) > 0$ on $(0,1)$. Therefore, either $\eta(u) > 0$ for all $u \in (0,1)$ or $\eta(u) < 0$ for all $u \in (0,1)$. However, the latter is excluded because by (1.47) we have $\int_0^1 \eta(u)\, du = 1$. Thus, $\eta(u) > 0$ on $(0,1)$, so that by (1.48), $\mathbf{1}(u \in B) - \mathbf{1}(u \in [0,1]\backslash B) = 1$ on $(0,1)$.

Consequently, we have the following theorem.

Theorem 1.20. *For every continuous density $h(u)$ on $(0,1)$ with support $(0,1)$ the sequence $\{\delta_m\}_{m=1}^{\infty}$ in Theorem 1.19 is unique, with*

$$\delta_m = \frac{\int_0^1 \sqrt{2}\cos(m\pi u)\sqrt{h(u)}\, du}{\int_0^1 \sqrt{h(u)}\, du}.$$

As is easy to verify, the same argument applies to the more general densities considered in Theorem 1.16:

Theorem 1.21. *Let the conditions of Theorem 1.16 be satisfied. In addition, let the density $w(x)$ and the orthonormal functions $\rho_m(x)$ be continuous on (a,b).*[19] *Then every continuous and positive density $f(x)$ on (a,b) has a unique series representation (1.36), with*

$$\delta_m = \frac{\int_a^b \rho_m(x)\sqrt{w(x)}\sqrt{f(x)}\,dx}{\int_0^1 \sqrt{w(x)}\sqrt{f(x)}\,dx}.$$

Moreover, note that Theorem 1.18 is a special case of Theorem 1.16. Therefore, the following corollary holds.

Corollary 1.3. *For every continuous and positive density $h(u)$ on $(0,1)$ the δ_m's in Theorem 1.18 are unique and given by*

$$\delta_m = (-1)^m \frac{\int_0^1 \sqrt{2}\cos(m\cdot\arccos(2u-1))\,(u(1-u))^{-1/4}\sqrt{h(u)}\,du}{\int_0^1 (u(1-u))^{-1/4}\sqrt{h(u)}\,du}.$$

1.6.5. Application to the MPH Competing Risks Model

Note that the distribution (1.6) in the MPH competing risks Weibull model (1.5) has density

$$h(u) = \int_0^\infty v^2 u^{v-1}\,dG(v),$$

which is obviously positive and continuous on $(0,1)$. However, if $G(1) > 0$, then $h(0) = \infty$; and if $E[V^2] = \int_0^\infty v^2\,dG(v) = \infty$, then $h(1) = \infty$. To allow for these possibilities, the series representation in Theorem 1.18 on the basis of Chebyshev polynomials seems an appropriate way of modeling $H(u)$ semi-nonparametrically, because then $h_n(0) = h_n(1) = \infty$ if $1 + \sqrt{2}\sum_{m=1}^\infty (-1)^m\delta_m \neq 0$ and $1 + \sqrt{2}\sum_{m=1}^\infty \delta_m \neq 0$. However, the approach in Theorem 1.19 is asymptotically applicable as well, because the condition $\sum_{m=1}^\infty \delta_m^2 < \infty$ does not preclude the possibilities that $\sum_{m=1}^\infty \delta_m = \infty$ and/or $\sum_{m=1}^\infty (-1)^m\delta_m = \infty$, which imply that $\lim_{n\to\infty} h_n(0) = \lim_{n\to\infty} h_n(1) = \infty$.

As said before, the actual log-likelihood in Bierens and Carvalho (2007) is more complicated than displayed in (1.7), due to right-censoring. In their case the log-likelihood involves the distribution function $H(u) = \int_0^\infty u^v\,dG(v)$ next to its density $h(u) = \int_0^\infty v u^{v-1}\,dG(v)$, where $h(1) = \int_0^\infty v\,dG(v) = 1$ due to the condition $E[V] = 1$. Note also that $G(1) > 0$ implies $h(0) = \infty$. Bierens and Carvalho (2007) use a series representation of $h(u)$ in terms of Legendre polynomials with SNP density $h_n(u)$ satisfying the restriction $h_n(1) = 1$. However, as argued in Section 1.6.1, the computation of the corresponding SNP distribution function $H_n(u)$ is complicated.

Due to the restriction $h_n(1) = 1$, the approach in Theorem 1.18 is not applicable as an alternative to the Legendre polynomial representation of $h(u) = \int_0^\infty v u^{v-1}\,dG(v)$, whereas the approach in Theorem 1.19 does not allow for $h_n(0) = \infty$. On the other hand, Bierens and Carvalho (2007) could have used $H_n(u) = \underline{H}_n(\sqrt{u})$, for example, where $\underline{H}_n$ is defined by (1.45), with density

$$h_n(u) = \frac{\left(1 + \sum_{m=1}^{n} \delta_m \sqrt{2} \cos\left(m\pi\sqrt{u}\right)\right)^2}{2\left(1 + \sum_{m=1}^{n} \delta_m^2\right)\sqrt{u}}$$

and δ_1 chosen such that

$$1 = \frac{\left(1 + \sqrt{2}\sum_{m=1}^{n}(-1)^m \delta_m\right)^2}{2\left(1 + \sum_{m=1}^{n} \delta_m^2\right)} \tag{1.51}$$

to enforce the restriction $h_n(1) = 1$.

1.6.6. Application to the First-Price Auction Model

In the first-price auction model, the value distribution $F(v)$ is defined on $(0,\infty)$, so at first sight a series expansion of the value density $f(v)$ in terms of Laguerre polynomials seems appropriate. However, any distribution function $F(v)$ on $(0,\infty)$ can be written as $F(v) = H(G(v))$, where $G(v)$ is an *a priori* chosen absolutely continuous distribution function with support $(0,\infty)$, so that $H(u) = F(G^{-1}(u))$ with density $h(u) = f((G^{-1}(u))/g(G^{-1}(u))$, where G^{-1} and g are the inverse and density of G, respectively. For example, choose $G(v) = 1 - \exp(-v)$, so that $g(v) = \exp(-v)$ and $G^{-1}(u) = \ln(1/(1-u))$.

The equilibrium bid function (1.8) can now be written as

$$\beta(v|H) = v - \frac{\int_{p_0}^{v} H(G(x))^{I-1}\, dx}{H(G(v))^{I-1}}, \qquad v \geq p_0. \tag{1.52}$$

Bierens and Song (2012) use the SNP approximation of $H(u)$ on the basis of Legendre polynomials, but using the results in Theorem 1.19 would have been much more convenient. In any case the integral in (1.52) has to be computed numerically.

Similarly, the conditional value distribution $F(v\exp(-\theta'X))$ in Bierens and Song (2013) can be written as $H(G(v\exp(-\theta'X)))$, where now H is modeled semi-nonparametrically according the results in Theorem 1.19. In this case the number of potential bidders $I = I(X)$ and the reservation price $p_0 = p_0(X)$ also depend on the auction-specific covariates X; but as shown in Bierens and Song (2013), $I(X)$ can be estimated nonparametrically and therefore may be treated as being observable, whereas $p_0(X)$ is directly observable. Then in the binding reservation price case the auction-specific equilibrium bid function becomes

$$\beta(v|H,\theta,X) = v - \frac{\int_{p_0(X)}^{v} H(G(x\cdot\exp(-\theta'X)))^{I(X)-1}\, dx}{H(G(v\exp(-\theta'X)))^{I(X)-1}}, \qquad v \geq p_0(X). \tag{1.53}$$

1.7. A Brief Review of Sieve Estimation

Recall from (1.30)–(1.32) that in the SNP index regression case the objective function takes the form

$$\widehat{Q}_N(\theta,\boldsymbol{\delta}_\infty) = \frac{1}{N}\sum_{j=1}^{N}\left(Y_j - \sum_{m=0}^{\infty}\delta_m\overline{p}_m(X_{1,j}+\theta' X_{2,j}|w_{\mathcal{N}[0,1]})\right)^2 K(||X_j||),$$

where $\boldsymbol{\delta}_\infty = (\delta_1,\delta_2,\delta_3,\dots)\in\mathbb{R}^\infty$ satisfies $\sum_{m=0}^{\infty}\delta_m^2<\infty$, with true parameters θ_0 and $\boldsymbol{\delta}_\infty^0 = (\delta_{0,1},\delta_{0,2},\delta_{0,3},\dots)$ satisfying

$$(\theta_0,\boldsymbol{\delta}_\infty^0) = \arg\min_{\theta,\boldsymbol{\delta}_\infty}\overline{Q}(\theta,\boldsymbol{\delta}_\infty) \tag{1.54}$$

subject to $\sum_{m=0}^{\infty}\delta_m^2<\infty$, where $\overline{Q}(\theta,\boldsymbol{\delta}_\infty) = E[\widehat{Q}_N(\theta,\boldsymbol{\delta}_\infty)]$.

Similarly, in the MPH competing risk model with $H(u)$ modeled semi-nonparametrically as, for example, $H(\sqrt{u}|\boldsymbol{\delta}_\infty) = \lim_{n\to\infty}\underline{H}_n(\sqrt{u})$ with $\underline{H}_n$ defined by (1.45), and subject to the restriction (1.51), the objective function is

$$\widehat{Q}_N(\theta,\boldsymbol{\delta}_\infty) = -\frac{1}{N}\ln\left(L_N(\alpha_1,\alpha_2,\beta_1,\beta_2,H(\sqrt{u}|\boldsymbol{\delta}_\infty))\right),$$
$$\theta = \left(\alpha_1',\alpha_2',\beta_1',\beta_2'\right)'.$$

with true parameters given by (1.54) with $\overline{Q}(\theta,\boldsymbol{\delta}_\infty) = E[\widehat{Q}_N(\theta,\boldsymbol{\delta}_\infty)]$.

In the first-price auction model with auction-specific covariates the function $\overline{Q}(\theta,\boldsymbol{\delta}_\infty)$ is the probability limit of the objective function $\widehat{Q}_N(\theta,\boldsymbol{\delta}_\infty)$ involved rather than the expectation. See Bierens and Song (2013).

Now let Θ be a compact parameter space for θ_0, and for each $n\geq 1$, let Δ_n be a compact space of nuisance parameters $\boldsymbol{\delta}_n = (\delta_1,\delta_2,\delta_3,\dots,\delta_n,0,0,0,\dots)$, endowed with metric $d(.,.)$, such that $\boldsymbol{\delta}_n^0 = (\delta_{0,1},\delta_{0,2},\delta_{0,3},\dots,\delta_{0,n},0,0,0,\dots)\in\Delta_n$. Note that $\boldsymbol{\delta}_\infty^0 \in \overline{\cup_{n=1}^{\infty}\Delta_n}$, where the bar denotes the closure.

The sieve estimator of $(\theta_0,\boldsymbol{\delta}_\infty^0)$ is defined as

$$\left(\widehat{\theta}_n,\widehat{\boldsymbol{\delta}}_n\right) = \arg\min_{(\theta,\boldsymbol{\delta}_n)\in\Theta\times\Delta_n}\widehat{Q}_N(\theta,\boldsymbol{\delta}_n).$$

Under some regularity conditions it can be shown that for an arbitrary subsequence n_N of the sample size N we obtain

$$p\lim_{N\to\infty}||\widehat{\theta}_{n_N}-\theta_0|| = 0 \quad\text{and}\quad p\lim_{N\to\infty}d\left(\widehat{\boldsymbol{\delta}}_{n_N},\boldsymbol{\delta}_\infty^0\right) = 0.$$

Moreover, under further regularity conditions the subsequence n_N can be chosen such that

$$\sqrt{N}(\widehat{\theta}_{n_N}-\theta_0)\xrightarrow{d} N[0,\Sigma].$$

See Shen (1997), Chen (2007), and Bierens (2013). As shown in Bierens (2013), the asymptotic variance matrix Σ can be estimated consistently by treating n_N as constant and then estimating the asymptotic variance matrix involved in the standard parametric way.

Note that Bierens and Carvalho (2007) assume that $\boldsymbol{\delta}_\infty^0 \in \cup_{n=1}^\infty \Delta_n$, so that for some n we have $\boldsymbol{\delta}_\infty^0 = \boldsymbol{\delta}_n^0 \in \Delta_n$. This is quite common in empirical applications because then the model is fully parametric, albeit with unknown dimension of the parameter space. See, for example, Gabler et al. (1993). The minimal order n in this case can be estimated consistently via an information criterion, such as the Hannan–Quinn (1979) and Schwarz (1978) information criteria. Asymptotically, the estimated order $\widehat{n}_N$ may then be treated as the true order, so that the consistency and asymptotic normality of the parameter estimates can be established in the standard parametric way.

In the case $\boldsymbol{\delta}_\infty^0 \in \overline{\cup_{n=1}^\infty \Delta_n} \setminus \cup_{n=1}^\infty \Delta_n$ the estimated sieve order $\widehat{n}_N$ via these information criteria will converge to ∞. Nevertheless, using $\widehat{n}_N$ in this case may preserve consistency of the sieve estimators, as in Bierens and Song (2012, Theorem 4), but whether asymptotic normality is also preserved is an open question.

1.8. Concluding Remarks

Admittedly, this discussion of the sieve estimation approach is very brief and incomplete. However, the main focus of this chapter is on SNP modeling. A full review of the sieve estimation approach is beyond the scope and size limitation of this chapter. Besides, a recent complete review has already been done by Chen (2007).

This chapter is part of the much wider area of *approximation theory*. The reader may wish to consult some textbooks on the latter—for example, Cheney (1982), Lorentz (1986), Powell (1981), and Rivlin (1981).

Notes

1. Of course, there are many more examples of SNP models.
2. See, for example, Bierens (2004, Theorem 3.10, p. 77).
3. See (1.41) below.
4. Note that due to the presence of scale parameters in the Weibull baseline hazards (1.3), the condition $E[V] = 1$ is merely a normalization of the condition that $E[V] < \infty$.
5. That is, $\{v : f(v) > 0\}$ is an interval.
6. See, for example, Bierens (2004, Theorem 7.A.1, p. 200).
7. Here the bar denotes the closure.
8. See, for example, Bierens (2004, Theorem 7.A.5, p. 202) for a proof of the projection theorem.

9. See, for example, Bierens (2004, Theorem 7.A.2., p. 200). The latter result is confined to the Hilbert space of zero-mean random variables with finite second moments, but its proof can easily be adapted to $\mathcal{R}$.
10. The existence of such a complete orthonormal sequence will be shown in the next section.
11. See, for example, Bierens (2004, Theorem 6.B.3, p. 168).
12. See, for example, Bierens (2004, Theorem 2.B.2, p. 168).
13. Charles Hermite (1822–1901).
14. Edmund Nicolas Laguerre (1834–1886).
15. Adrien-Marie Legendre (1752–1833).
16. See, for example, Bierens (1994, Theorem 3.1.1, p. 50).
17. See Section 1.7.
18. See, for example, Young (1988, Chapter 5).
19. The latter is the case if we choose $\rho_m(x) = \overline{p}(x|w)$.

References

Abbring, Jaap H., and Gerard J. van den Berg. 2003. "The Identifiability of the Mixed Proportional Hazards Competing Risks Model." *Journal of the Royal Statistical Society B*, **65**, pp. 701–710.

Bierens, Herman J. 1994. *Topics in Advanced Econometrics*: *Estimation, Testing, and Specification of Cross-Section and Time Series Models*. Cambridge, UK: Cambridge University Press.

Bierens, Herman J. 2004. *Introduction to the Mathematical and Statistical Foundations of Econometrics*. Cambridge, UK: Cambridge University Press.

Bierens, Herman J. 2008. "Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results." *Econometric Theory*, **24**, pp. 749–794.

Bierens, Herman J. 2013. "Consistency and Asymptotic Normality of ML Sieve Estimators Under Low-Level Conditions." Forthcoming in *Econometric Theory*. (http://econ.la.psu.edu/hbierens/SNPMODELS.PDF).

Bierens, Herman J., and Hosin Song. 2012. "Semi-Nonparametric Estimation of Independently and Identically Repeated First-Price Auctions via an Integrated Simulated Moments Method." *Journal of Econometrics*, **168**, pp. 108–119.

Bierens, Herman J., and Hosin Song. 2013. "Semi-Nonparametric Modeling and Estimation of First-Price Auction Models with Auction-Specific Heterogeneity." Working paper (http://econ.la.psu.edu/hbierens/AUCTIONS_HETERO.PDF).

Bierens, Herman J., and Jose R. Carvalho. 2006. "Separate Appendix to: Semi-Nonparametric Competing Risks Analysis of Recidivism" (http://econ.la.psu.edu/~hbierens/RECIDIVISM_APP.PDF).

Bierens, Herman J., and Jose R. Carvalho. 2007. "Semi-Nonparametric Competing Risks Analysis of Recidivism." *Journal of Applied Econometrics*, **22**, pp. 971–993.

Chen, Xiaohong. 2007. "Large Sample Sieve Estimation of Semi-Nonparametric Models." Chapter 76 in *Handbook of Econometrics*, Vol. 6B, eds. James J. Heckman and Edward E. Leamer. Amsterdam: North-Holland.

Cheney, Elliott W. 1982. *Introduction to Approximation Theory*. Providence, RI: AMS Chelsea Publishers.

Eastwood, Brian J., and A. Ronald Gallant. 1991. "Adaptive Rules for Semi-Nonparametric Estimators that Achieve Asymptotic Normality." *Econometric Theory,* **7**, pp. 307–340.

Elbers, Chris, and Geert Ridder. 1982. "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model." *Review of Economic Studies*, **49**, pp. 403–409.

Gabler, Siegfried, Francois Laisney and Michael Lechner. 1993. "Seminonparametric Estimation of Binary-Choice Models with an Application to Labor-Force Participation." *Journal of Business & Economic Statistics,* **11**, pp. 61–80.

Gallant, A. Ronald. 1981. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form." *Journal of Econometrics,* **15**, pp. 211–245.

Gallant, A. Ronald, and Douglas W. Nychka. 1987. "Semi-Nonparametric Maximum Likelihood Estimation." *Econometrica,* **55**, pp. 363–390.

Grenander, Ulf. 1981. *Abstract Inference.* New York: John Wiley.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong. 2000. "Optimal Nonparametric Estimation of First-Price Auctions." *Econometrica,* **68**, pp. 525–574.

Hahn, Jinyong. 1994. "The Efficiency Bound of the Mixed Proportional Hazard Model." *Review of Economic Studies,* **61**, pp. 607–629.

Hamming, Richard W. 1973. *Numerical Methods for Scientists and Engineers.* New York: Dover Publications.

Hannan, Edward J., and Barry G. Quinn. 1979. "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society, Series B,* **41**, pp. 190–195.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica,* **47**, pp. 153–161.

Heckman, James J., and Bo E. Honore. 1989. "The Identifiability of the Competing Risks Model." *Biometrika,* **76**, pp. 325–330.

Heckman, James J., and Burton Singer. 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica,* **52**, pp. 271–320.

Jennrich, Robert I. 1969. "Asymptotic Properties of Nonlinear Least Squares Estimators." *Annals of Mathematical Statistics,* **40**, pp. 633–643.

Krishna, Vijay. 2002. *Auction Theory.* San Diego: Academic Press.

Kronmal, R., and M. Tarter. 1968. "The Estimation of Densities and Cumulatives by Fourier Series Methods." *Journal of the American Statistical Association,* **63**, pp. 925–952.

Lancaster, Tony. 1979. "Econometric Methods for the Duration of Unemployment." *Econometrica,* **47**, pp. 939–956.

Lorentz, George G. 1986. *Approximation of Functions.* Providence, RI: AMS Chelsea Publishers.

Newey, Whitney K. 1990. "Semiparametric Efficiency Bounds." *Journal of Applied Econometrics*, **5**, pp. 99–135.

Newey, Whitney K. 1997. "Convergence Rates and Asymptotic Normality for Series Estimators." *Journal of Econometrics,* **79**, pp. 147–168.

Powell, Michael J. D. 1981. *Approximation Theory and Methods.* Cambridge, U.K.: Cambridge University Press.

Ridder, Geert, and Tiemen Woutersen. 2003. "The Singularity of the Efficiency Bound of the Mixed Proportional Hazard Model." *Econometrica,* **71**, pp. 1579–1589.

Riley, John G., and William F. Samuelson. 1981. "Optimal Auctions." *American Economic Review,* **71**, pp. 381–392.

Rivlin, Theodore J. 1981. *An Introduction to the Approximation of Functions*. New York: Dover Publications.

Schwarz, Gideon. 1978. “Estimating the Dimension of a Model.” *Annals of Statistics,* **6**, pp. 461–464.

Shen, Xiaotong. 1997. “On the Method of Sieves and Penalization.” *Annals of Statistics,* **25**, pp. 2555–2591.

Stewart, Mark B. 2004. “Semi-Nonparametric Estimation of Extended Ordered Probit Models.” *The Stata Journal,* **4**, pp. 27–39.

van den Berg, Gerard. 2000. “Duration models: Specification, identification, and multiple duration.” Chapter 55 in *Handbook of Econometrics*, Vol. 5, eds. James J. Heckman and Edward E. Leamer. Amsterdam: North-Holland.

van der Vaart, Aart W. 1998. *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press.

Wald, Abraham. 1949. “Note on the Consistency of the Maximum Likelihood Estimate.” *Annals of Mathematical Statistics,* **20**, pp. 595–601.

Wold, Herman. 1938. *A Study in the Analysis of Stationary Time Series*. Stockholm: Almqvist and Wiksell.

Young, Nicholas. 1988. *An Introduction to Hilbert Space*. Cambridge, UK: Cambridge University Press.