Hilbert Space Theory and Its Applications to Semi-Nonparametric Modeling and Inference

Herman J. Bierens

Pennsylvania State University

Current version: January 25, 2012

Chapter 1

Introduction

As is well known, every vector in a Euclidean space can be represented as a linear combination of orthonormal vectors. Similarly, using Hilbert space theory, we can represent certain classes of Borel measurable functions[1] by countable infinite linear combinations of orthonormal functions, which allows us to approximate these functions arbitrarily close by finite linear combinations of these orthonormal functions. This is the basis for semi-nonparametric (SNP) modeling, where only a part of the model involved is parametrized, and the non-specified part is an unknown function which is approximated by a series expansion. See for example Chen (2007) for a recent survey, and Bickel et al (1998). There is also a substantial literature on estimation of semi-nonparametric models using nonparametric kernel density and/or regression estimators (see for example Horowitz 1998), but these approaches are beyond our scope.

Gallant (1981) was the first econometrician to proposed Fourier series expansions as a way to model unknown functions. Gallant's approach is actually nonparametric in that no Euclidean parameters are involved. 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 estimation of Heckman's (1979) sample selection model, where the bivariate error distribution of the latent

[1] See for example Bierens (2004, Ch. 2) for the definition of Borel measurability of functions.

variable equations is modeled semi-nonparametrically using an Hermite expansion of the error density.

Another example of a semi-nonparametric model is the mixed proportional hazard (MPH) model proposed by Lancaster (1979). In this model the hazard function is the product of three factors, the baseline hazard which depends only on the duration, the systematic hazard which is a function of the observable covariates, and an unobserved non-negative random variable representing neglected heterogeneity. Elbers and Ridder (1982) have shown that under some mild conditions and normalizations the MPH model is nonparametrically identified. Heckman and Singer (1984) propose to estimate 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.

In chapter 2 I will explain what a Hilbert space is, and provide examples of non-Euclidean Hilbert spaces, in particular Hilbert spaces of Borel measurable functions and random variables. In chapter 3 I will discuss projections on sub-Hilbert spaces and their properties. One of the results involved is the famous Wold (1938) decomposition theorem, which will be derived first in general terms and then for covariance stationary time series. Also, the fundamental role of the Wold decomposition in time series analysis and empirical macro-econometrics will be pointed out.

The main focus of this book, however, is on Hilbert spaces of square integrable Borel measurable real functions and the various orthonormal sequences that span these Hilbert spaces, as the basis for semi-nonparametric modeling and inference. Therefore, following Hamming (1973), in chapter 4 I will review the various ways one can construct orthonormal polynomials that span a given Hilbert space of functions. In chapter 5 I will show that any square integrable Borel measurable real function on the unit interval can be written as a linear combination of the cosine series $\{\cos(k\pi u)\}_{k=0}^{\infty}$, $u \in [0,1]$. This result is related to classical Fourier analysis, which will also be reviewed. The significance of this result is that it yields closed form series representations of arbitrary density and distribution functions, as will be shown in chapter 6, whereas in the approach of Gallant and Nychka (1987), which is based on Hermite polynomials, and the approach of Bierens (2008) and Bierens and Carvalho (2007), which is based on Legendre polynomials, the computation of their density and distribution functions has to be done iteratively. In chapter 7 I will show how to construct compact metric spaces of density and distribution functions based on the cosine series expansion.

The applications to semi-nonparametric models, based on Bierens. (2011), will be added to this manuscript in due course.

Throughout this manuscript the set of positive integers will be denoted by $\mathbb{N}$, and the set of non-negative integers by $\mathbb{N}_0$. Moreover, the well-known indicator function will be denoted by $I(.)$, and $\mathbf{i} = \sqrt{-1}$.

Part I

Hilbert spaces

Chapter 2

Introduction to Hilbert spaces

In this chapter I will review the concepts of vector spaces, inner products and Cauchy sequences, and provide examples of Hilbert spaces.

2.1 Vector spaces

The notion of a vector space should be known from linear algebra:

Definition 2.1. *Let $\mathcal{V}$ be a set endowed with two operations, the operation "addition", denoted by "+", which maps each pair (x, y) in $\mathcal{V} \times \mathcal{V}$ into $\mathcal{V}$, and the operation "scalar multiplication", denoted by a dot (.), which maps each pair (c, x) in $\mathbb{R} \times \mathcal{V}$ [or $\mathbb{C} \times \mathcal{V}$] into $\mathcal{V}$. Thus, a scalar is a real or complex number. The set $\mathcal{V}$ is called a real [complex] vector space if the addition and multiplication operations involved satisfy the following rules, for all x, y and z* in $\mathcal{V}$, *and all scalars c, c_1 and c_2 in* $\mathbb{R}$ [$\mathbb{C}$]:
(a) $x + y = y + x$;
(b) $x + (y + z) = (x + y) + z$;
(c) *There is a unique zero vector* 0 *in* $\mathcal{V}$ *such that* $x + 0 = x$;
(d) For each x there exists a unique vector $-x$ in $\mathcal{V}$ such that $x + (-x) = 0$;[1]
(e) $1.x = x$;
(f) $(c_1c_2).x = c_1.(c_2.x)$;
(g) $c.(x + y) = c.x + c.y$;
(h) $(c_1 + c_2).x = c_1.x + c_2.x$.

[1] Also denoted by $x - x = 0$.

It is trivial to verify that the Euclidean space $\mathbb{R}^n$ is a real vector space. However, the notion of a vector space is much more general. For example, let $\mathcal{V}$ be the space of all continuous functions on $\mathbb{R}^n$, with pointwise addition and scalar multiplication defined the same way as for real numbers. Then it is easy to verify that this space is a real vector space.

Another (but weird) example of a vector space is the space $\mathcal{V}$ of positive real numbers endowed with the "addition" operation $x + y = x.y$ and the "scalar multiplication" $c.x = x^c$. In this case the null vector 0 is the number 1, and $-x = 1/x$.

Definition 2.2. *A subspace $\mathcal{V}_0$ of a vector space $\mathcal{V}$ is a non-empty subset of $\mathcal{V}$ which satisfies the following two requirements*:
(*a*) *For any pair x, y in $\mathcal{V}_0$, $x + y$ is in $\mathcal{V}_0$*;
(*b*) *For any x* in *$\mathcal{V}_0$ and any scalar c, $c.x$ is in $\mathcal{V}_0$.*

Thus, a subspace $\mathcal{V}_0$ of a vector space is closed under linear combinations: any linear combination of elements in $\mathcal{V}_0$ is an element of $\mathcal{V}_0$.

It is not hard to verify that a subspace of a vector space is a vector space itself, because the rules (a) through (h) in Definition 2.1 are inherited from the "host" vector space $\mathcal{V}$. In particular, any subspace contains the null vector 0, as follows from part (b) of Definition 2.2 with $c = 0$.

2.2 Inner product and norm

As is well-known, in a Euclidean space $\mathbb{R}^n$ the inner product of a pair of vectors x and y is defined as $x'y$, which is a mapping $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ with the following properties:
(a) $x'y = y'x$,
(b) $(cx)'y = c(x'y)$ for arbitrary $c \in \mathbb{R}$,
(c) $(x + y)'z = x'z + y'z$,
(d) $x'x > 0$ if and only if $x \neq 0$.
Moreover, the norm of a vector $x \in \mathbb{R}^n$ is defined as $||x|| = \sqrt{x'x}$. Of course, in $\mathbb{R}$ the inner product is the ordinary product $x.y$.

Mimicking these four properties, we can define more general inner products with associated norms as follows.

Definition 2.3. *An inner product on a real vector space* $\mathcal{V}$ *is a real function* $\langle x, y\rangle$: $\mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ *such that for all* x, y, z *in* $\mathcal{V}$ *and all* c *in* $\mathbb{R}$,
(*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 $x \neq 0$.
An inner product on a complex vector space is defined similarly. The inner product is then complex-valued, $\langle x, y\rangle$: $\mathcal{V} \times \mathcal{V} \rightarrow \mathbb{C}$. *Condition* (*1*) *then becomes*
(*1**) $\langle x, y\rangle = \overline{\langle y, x\rangle}$,[2]
and (*2*) *now holds for all complex and real numbers* c. *Note that also in this case* $\langle x, x\rangle$ *is real valued.*[3] *A vector space endowed with an inner product is called an inner product space. Finally, the norm of* x *in* $\mathcal{V}$ *is defined as* $||x|| = \sqrt{\langle x, x\rangle}$

For example, in the vector space $C[0, 1]$ of continuous real functions on $[0, 1]$, the integral $\langle f, g\rangle = \int_0^1 f(u)\,g(u)\,du$ is an inner product, with norm $\|f\| = \sqrt{\int_0^1 f(u)^2\,du}$. Moreover, in the vector space of zero-mean random variables with finite second moments the covariance $\langle X, Y\rangle = E[X.Y]$ is an inner product, with norm $\|X\| = \sqrt{E[X^2]}$.

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

Theorem 2.1. (*Cauchy-Schwarz inequality*) $|\langle x, y\rangle| \leq ||x||.||y||$.

Given the norm $||x|| = \sqrt{\langle x, x\rangle}$, the following properties hold:

$$||x|| > 0 \text{ if } x \neq 0; \tag{2.1}$$
$$||c.x|| = |c|.||x||; \tag{2.2}$$
$$||x + y|| \leq ||x|| + ||y||. \tag{2.3}$$

The latter is known as the triangular inequality.

[2] The bar denotes the complex conjugate: for $z = a + i.b$, $\overline{z} = a - i.b$.
[3] Because $\langle x, x\rangle = \overline{\langle x, x\rangle}$ implies that $\langle x, x\rangle \in \mathbb{R}$.

The properties (2.1) and (2.2) follow trivially from Definition 2.3. In the case of a real vector space the triangular inequality (2.3) follows from

$$\begin{aligned} ||x+y||^2 &= \langle x+y, x+y\rangle = \langle x,x\rangle + 2\langle x,y\rangle + \langle y,y\rangle \\ &= ||x||^2 + 2\langle x,y\rangle + ||y||^2 \le ||x||^2 + 2|\langle x,y\rangle| + ||y||^2 \\ &\le ||x||^2 + 2||x||.||y|| + ||y||^2 = (||x|| + ||y||)^2 \end{aligned}$$

where the last inequality is due to Theorem 2.1.

In a Euclidean space, a pair x, y of vectors is orthogonal if $x'y = 0$, and orthonormal if also $||x|| = ||y|| = 1$. Similarly,

Definition 2.4. *Elements x and y in a inner product space with associated norm are orthogonal if $\langle x, y\rangle = 0$, which is also denoted by $x \perp y$, and are orthonormal if in addition $||x|| = ||y|| = 1$.*

A norm can also be defined directly:

Definition 2.5. *A norm on a vector space $\mathcal{V}$ is a mapping $||.||\colon \mathcal{V} \to [0, \infty)$ such that for all x and y in $\mathcal{V}$ and all scalars c the properties* (2.1), (2.2) *and* (2.3) *hold. A vector space endowed with a norm is called a normed space.*

2.3 Metric spaces

A norm $||.||$ defines a metric $d(x, y) = ||x - y||$ on $\mathcal{V}$, i.e., a function that measures the distance between two elements x and y of $\mathcal{V}$, for which (trivially) the following four properties hold. For all x, y and z in $\mathcal{V}$,

$$d(x,y) = d(y,x) \tag{2.4}$$

$$d(x,y) > 0 \; if \; x \ne y; \tag{2.5}$$

$$d(x,x) = 0; \tag{2.6}$$

$$d(x,z) \le d(x,y) + d(y,z). \tag{2.7}$$

Again, the property (2.7) is known as the triangular inequality.

A metric can also be defined directly:

Definition 2.6. *A metric on a space $\mathcal{M}$ is a mapping $d(.,.)\colon \mathcal{M} \times \mathcal{M} \to [0, \infty)$ satisfying the properties* (2.4) *through* (2.7) *for all x, y and z in $\mathcal{M}$. A space endowed with a metric is called a metric space.*

In this definition the space $\mathcal{M}$ is not necessarily a vector space: Any space endowed with a metric is a metric space. For example, let $\mathcal{M}$ be the space of density functions on $[0,1]$, endowed with the metric

$$d(f,g) = \int_0^1 \left(\sqrt{f(u)} - \sqrt{g(u)}\right)^2 du.$$

This space is not a vector space, and it is not possible to define an inner product on it.

2.4 Convergence of Cauchy sequences

A vector space $\mathcal{V}$ endowed with an inner product $\langle x,y\rangle$ and associated norm $||x|| = \sqrt{\langle x,y\rangle}$ and metric $||x-y||$ is called a *pre-Hilbert space*. The reason for the "pre" is that a fundamental property is still missing, namely that every Cauchy sequence has a limit in $\mathcal{V}$.

Definition 2.7. *A sequence of elements x_n of a metric space with metric $d(.,.)$ is called a Cauchy sequence if for every $\varepsilon > 0$ there exists an $n_0(\varepsilon)$ such that for all $k,m \geq n_0(\varepsilon)$, $d(x_k,x_m) < \varepsilon$.*

For example, in the Euclidean space $\mathbb{R}^p$ with finite dimension p every Cauchy sequence converges to a limit in $\mathbb{R}^p$, and the same applies to the space $\mathbb{C}^p$ of p-dimensional vectors with complex-valued components, endowed with the inner product

$$\begin{aligned} \langle x,y\rangle &= \overline{x}'y = (\mathrm{Re}\,(x) - \mathbf{i}.\,\mathrm{Im}\,(x))'\,(\mathrm{Re}(y) + \mathbf{i}.\,\mathrm{Im}(y)) \\ &= \left(\mathrm{Re}\,(x)'\,\mathrm{Re}(y) + \mathrm{Im}\,(x)'\,\mathrm{Im}(y)\right) \\ &\quad + \mathbf{i}.\,(\mathrm{Re}(x)'\,\mathrm{Im}(y) - \mathrm{Im}(x)'\,\mathrm{Re}(y)) \end{aligned} \tag{2.8}$$

and associated norm and metric. It is an easy exercise to check that (2.8) satisfies the conditions in Definition 2.3. Thus,

Theorem 2.2. *Every Cauchy sequence in $\mathbb{R}^p$ or $\mathbb{C}^p$ has a limit in that space.*

To demonstrate the role of the Cauchy convergence property, consider the space $C[0,1]$ of continuous real functions on $[0,1]$, i.e., each $f \in C[0,1]$ is continuous on $(0,1)$, and $f(0) = \lim_{u\downarrow 0} f(u)$ and $f(1) = \lim_{u\uparrow 1} f(u)$ are finite. 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 $C[0,1]$:

$$\begin{aligned} f_n(u) &= \begin{cases} 0 & \text{for } 0 \le u < 0.5 \\ 2^n(u-0.5) & \text{for } 0.5 \le u < 0.5 + 2^{-n} \\ 1 & \text{for } 0.5 + 2^{-n} \le u \le 1, \end{cases} \\ n &= 1,2,3,..... \end{aligned}$$

It is an easy calculus exercise to verify that $||f_k - f_m||^2 = \int_0^1 (f_k(u) - f_m(u))^2 du < \frac{1}{3}\left(2^{-k} + 2^{-m}\right)$, hence f_n is a Cauchy sequence in $C[0,1]$. Moreover, if follows from the bounded convergence theorem that $\lim_{n\to\infty} ||f_n - f|| = 0$, where $f(u) = I(u > 0.5)$. 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 closed under convergence.

2.5 Hilbert spaces and sub-Hilbert spaces

2.5.1 Hilbert spaces versus Banach spaces

It is usually quite easy to define an inner product on a vector space, and the same vector space can often be endowed with different inner products. For example, for the space of square integrable Borel measurable functions on $[0,1]$ we can define an inner product by $\langle f,g\rangle = \int_0^1 f(u)g(u)du$ but also by $\langle f,g\rangle = \int_0^1 uf(u)g(u)du$, for example. However, to make such a space a Hilbert space the inner product must be chosen such that every Cauchy sequence converges to a limit in that space. The requirement that every Cauchy sequence in a Hilbert space has a limit in that space makes a Hilbert space closed under convergence, which generates all kinds of useful properties, similar to Euclidean spaces.

Definition 2.8. *A Hilbert space $\mathcal{H}$ is a vector space endowed with an inner product and associated norm and metric such that every Cauchy sequence in $\mathcal{H}$ has a limit in $\mathcal{H}$. The way the inner product $\langle x,y\rangle$ is defined, together with the associated norm and metric, will be called the topology of $\mathcal{H}$.*

Note that the limit of a Cauchy sequence in a Hilbert space is unique. To see this, suppose that a Cauchy sequence $x_n \in \mathcal{H}$ has two limits: $\lim_{n\to\infty} ||x_n - x|| = 0$ and $\lim_{n\to\infty} ||x_n - x_*|| = 0$. Then $\|x_* - x\| = \|x_* - x_n + x_n - x\| \leq ||x_n - x|| + ||x_n - x_*|| \to 0$ as $n \to \infty$. Conversely, any convergent sequence in a Hilbert space is a Cauchy sequence, because $\lim_{n\to\infty} ||x_n - x|| = 0$ implies that $||x_k - x_m|| = ||(x_k - x) + (x - x_m)|| \leq ||x_k - x|| + ||x_m - x|| \to 0$ as $\min(k, m) \to \infty$.

In Definition 2.5 the norm $||.||$ was defined directly, without reference to an inner product, giving rise to the definition of a normed space $\mathcal{N}$, for example. If we endow $\mathcal{N}$ with the metric $||x - y||$ and require that every Cauchy sequence in $\mathcal{N}$ has a limit in $\mathcal{N}$ then the space $\mathcal{N}$ becomes a Banach space. The difference between a Hilbert space and a Banach space is the source of the norm: In an Hilbert space the norm is defined on the basis of an inner product whereas in the case of a Banach space the norm is defined directly. Consequently, in a Banach space the notion of inner product is nonexisting, and so is the notion of orthogonality.

2.5.2 Linear manifolds and sub-Hilbert spaces

Because a Hilbert space is a vector space, we can define a subspace of a Hilbert space in the same way as for vector spaces (see Definition 2.2), and endow it with the same inner product, norm and metric as Hilbert space involved. Such a subspace is called a linear manifold:

Definition 2.9. *A linear manifold $\mathcal{M}$ of a Hilbert space $\mathcal{H}$ is a subspace of $\mathcal{H}$ endowed with the topology of $\mathcal{H}$.*

However, a linear manifold $\mathcal{M}$ is not necessarily a Hilbert space itself. In general there is no guarantee that every Cauchy sequence in $\mathcal{M}$ takes a limit in $\mathcal{M}$. If so the linear manifold $\mathcal{M}$ needs to be extended by augmenting it with the limits of all Cauchy sequence in $\mathcal{M}$. The resulting extended linear manifold coincides with the closure $\overline{\mathcal{M}}$ of $\mathcal{M}$. Recall that $\mathcal{M}$ is a subset of the metric space $\mathcal{H}$, and that a point of closure of $\mathcal{M}$ is an element $\overline{x}$ such that for each $\varepsilon > 0$ there exists a $z \in \mathcal{M}$ and a $y \in \mathcal{H}\backslash\mathcal{M}$ such that $||\overline{x} - z|| < \varepsilon$ and $||\overline{x} - y|| < \varepsilon$. The set of all points of closure of $\mathcal{M}$ is called the border of $\mathcal{M}$, denoted by $\partial\mathcal{M}$, and the closure of $\mathcal{M}$, denoted by $\overline{\mathcal{M}}$, is the union of $\mathcal{M}$ and its border: $\overline{\mathcal{M}} = \mathcal{M} \cup \partial\mathcal{M}$.

Theorem 2.3. *The closure $\overline{\mathcal{M}}$ of a linear manifold $\mathcal{M}$ is a Hilbert space.*

In other words, $\overline{\mathcal{M}}$ is a sub-Hilbert space.

2.5.3 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 $\mathcal{H}$. Let $\mathcal{M}_m$ be the linear manifold spanned by $x_1, ..., x_m$, i.e., $\mathcal{M}_m$ consists of all linear combinations of $x_1, ..., x_m$. Then it follows similar to the proof of Theorem 2.3 that

Lemma 2.1. *$\mathcal{M}_m$ is a Hilbert space.*

Definition 2.10. *The space $\mathcal{M}_\infty = \overline{\cup_{n=1}^{\infty}\mathcal{M}_n}$ is called the space spanned by $\{x_j\}_{j=1}^{\infty}$, and is also denoted by* $\text{span}(\{x_j\}_{j=1}^{\infty})$.

It follows similar to the proof of Theorem 2.3 that

Lemma 2.2. *$\mathcal{M}_\infty$ is a Hilbert space.*

Remark. In the sequel a sub-Hilbert space will be referred to as a "subspace".

Definition 2.11. *A sequence $\{x_k\}_{k=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ is called complete if $\mathcal{H} = \text{span}(\{x_j\}_{j=1}^{\infty})$.*

2.6 Examples of non-Euclidean Hilbert spaces

2.6.1 A Hilbert space of random variables

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

Theorem 2.4. *The space $\mathcal{R}$ is a Hilbert space.*

2.6.2 Hilbert spaces of functions

Let $w(x)$ be a probability density on $\mathbb{R}$ and let $L^2(w)$ be the space of Borel measurable real functions f on $\mathbb{R}$ satisfying

$$\int_{-\infty}^{\infty} f(x)^2 w(x)dx < \infty$$

where the integral is the Lebesgue integral, endowed with the inner product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g(x)w(x)dx$$

and associated norm $||f|| = \sqrt{\langle f, f\rangle} = \sqrt{\int_{-\infty}^{\infty} f(x)^2 w(x)dx}$ and metric $||f - g||$. Then for $f, g \in L^2(w)$, $\langle f, g\rangle = E[f(X)g(X)]$, where X is a random drawing from the distribution with density $w(x)$, hence it follows from Theorem 2.4 that $L^2(w)$ is a Hilbert space.

2.7 Appendix: Proofs

2.7.1 Theorem 2.1

Let the vector space involved be complex. It follows from the properties (1)-(4) in Definition 2.3 that for any complex valued λ,

$$\begin{aligned}
0 &\leq \langle x + \lambda y, x + \lambda y\rangle \\
&= \langle x, x\rangle + \langle \lambda y, x\rangle + \langle x, \lambda y\rangle + \langle \lambda y, \lambda y\rangle \\
&= ||x||^2 + \lambda \langle y, x\rangle + \overline{\langle \lambda y, x\rangle} + \lambda \langle y, \lambda y\rangle \\
&= ||x||^2 + \lambda\overline{\langle x, y\rangle} + \overline{\lambda \langle y, x\rangle} + \lambda\overline{\langle \lambda y, y\rangle} \\
&= ||x||^2 + \lambda\overline{\langle x, y\rangle} + \overline{\lambda}.\overline{\langle y, x\rangle} + \lambda.\overline{\lambda \langle y, y\rangle} \\
&= ||x||^2 + \lambda\overline{\langle x, y\rangle} + \overline{\lambda}.\langle x, y\rangle + \lambda.\overline{\lambda} \langle y, y\rangle \\
&= ||x||^2 + \lambda\overline{\langle x, y\rangle} + \overline{\lambda}.\langle x, y\rangle + \lambda.\overline{\lambda}||y||^2
\end{aligned}$$

Next, note that

$$\begin{aligned}
\lambda\overline{\langle x, y\rangle} + \overline{\lambda}.\langle x, y\rangle &= (\mathrm{Re}(\lambda) + \mathbf{i}.\mathrm{Im}(\lambda))(\mathrm{Re}(\langle x, y\rangle) - \mathbf{i}.\mathrm{Im}(\langle x, y\rangle)) \\
&\quad + (\mathrm{Re}(\lambda) - \mathbf{i}.\mathrm{Im}(\lambda))(\mathrm{Re}(\langle x, y\rangle) + \mathbf{i}.\mathrm{Im}(\langle x, y\rangle)) \\
&= 2(\mathrm{Re}(\lambda)\mathrm{Re}(\langle x, y\rangle) + \mathrm{Im}(\lambda)\mathrm{Im}(\langle x, y\rangle))
\end{aligned}$$

and

$$\begin{aligned}\lambda.\overline{\lambda} &= (\mathrm{Re}\,(\lambda) + \mathbf{i}.\,\mathrm{Im}\,(\lambda))\,(\mathrm{Re}\,(\lambda) - \mathbf{i}.\,\mathrm{Im}\,(\lambda)) \\ &= (\mathrm{Re}\,(\lambda))^2 + (\mathrm{Im}\,(\lambda))^2\end{aligned}$$

Hence

$$\begin{aligned}0 \leq\ & ||x||^2 + 2\,(\mathrm{Re}\,(\lambda)\,\mathrm{Re}\,(\langle x, y\rangle) + \mathrm{Im}\,(\lambda)\,\mathrm{Im}\,(\langle x, y\rangle)) \\ & + \left((\mathrm{Re}\,(\lambda))^2 + (\mathrm{Im}\,(\lambda))^2\right).||y||^2\end{aligned} \tag{2.9}$$

The latter is minimal for

$$\mathrm{Re}\,(\lambda) = -\frac{\mathrm{Re}\,(\langle x, y\rangle)}{||y||^2},\ \mathrm{Im}\,(\lambda) = -\frac{\mathrm{Im}\,(\langle x, y\rangle)}{||y||^2}.$$

Substituting these solutions in (2.9) yields

$$0 \leq ||x||^2 - \frac{1}{||y||^2}\left((\mathrm{Re}\,(\langle x, y\rangle))^2 + (\mathrm{Im}\,(\langle x, y\rangle))^2\right) = ||x||^2 - \frac{|\langle x, y\rangle|^2}{||y||^2}$$

and thus $|\langle x, y\rangle| \leq ||x||.||y||$.

2.7.2 Theorem 2.2

Let x_n be a Cauchy sequence in $\mathbb{R}$, and denote $\overline{x} = \limsup_{n\to\infty} x_n$. Let us first show that $\overline{x} < \infty$, as follows. By the definition of "limsup" there exists a subsequence n_k such that $\overline{x} = \lim_{k\to\infty} x_{n_k}$. Note that this x_{n_k} is also a Cauchy sequence, hence for arbitrary $\varepsilon > 0$ there exists an index k_0 such that $|x_{n_k} - x_{n_m}| < \varepsilon$ for all $k, m \geq k_0$. Keeping $m \geq k_0$ fixed and letting $k \to \infty$ it follows that $|\overline{x} - x_{n_m}| < \varepsilon$, hence $\overline{x} < \infty$. By a similar argument it follows that $\underline{x} = \liminf_{n\to\infty} x_n > -\infty$. Thus, we can find an index k_0 and subsequences $n_{1,k}$ and $n_{2,m}$ such that for all $k, m \geq k_0$, $\left|\overline{x} - x_{n_{1,m}}\right| < \varepsilon$, $\left|\underline{x} - x_{n_{2,m}}\right| < \varepsilon$ and $\left|x_{n_{1,m}} - x_{n_{2,m}}\right| < \varepsilon$, hence $|\overline{x} - \underline{x}| < 3\varepsilon$. Since ε was arbitrary, it follows now that $\overline{x} = \underline{x} = x$, which implies that $\lim_{n\to\infty} x_n = x \in \mathbb{R}$. By applying this argument to the real and imaginary parts of a complex valued Cauchy sequence x_n it follows that $\lim_{n\to\infty} x_n = x \in \mathbb{C}$. Moreover, applying this argument to each component of a (complex) vector valued Cauchy sequence the results for the cases $\mathbb{R}^p$ and $\mathbb{C}^p$ follow.

2.7.3 Theorem 2.3

Let x_n be a Cauchy sequence in $\overline{\mathcal{M}} \subset \mathcal{H}$. Then x_n has a limit $\overline{x} \in \mathcal{H}$, i.e., $\lim_{n\to\infty} ||x_n - \overline{x}|| = 0$. Suppose that $\overline{x} \notin \overline{\mathcal{M}}$. Since $\overline{\mathcal{M}}$ 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 $\overline{\mathcal{M}}$: $\mathcal{N}(\overline{x}, \varepsilon) \cap \overline{\mathcal{M}} = \emptyset$. 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 $x_n \notin \overline{\mathcal{M}}$ for all $n > \underline{n}(\varepsilon)$, which contradicts $x_n \in \overline{\mathcal{M}}$ for all n.

2.7.4 Lemma 2.1

Without loss of generality we may assume that the $m \times m$ matrix X_m with elements $\langle x_i, x_j \rangle$ is nonsingular, as otherwise we can re-arrange the x_j's such that $\mathcal{M}_m = \mathcal{M}_r$ with $r = \text{rank}(X_m)$. Let $y_{n,m} = \sum_{j=1}^m c_{j,n} x_j$ be a Cauchy sequence in $\mathcal{M}_m$. Then

$$\begin{aligned} ||y_{n_1,m} - y_{n_2,m}||^2 &= \left\| \sum_{j=1}^m (c_{j,n_1} - c_{j,n_2}) x_j \right\|^2 \\ &= \sum_{i=1}^m \sum_{j=1}^m (c_{i,n_1} - c_{i,n_2})(c_{j,n_1} - c_{j,n_2}) \langle x_i, x_j \rangle \to 0 \end{aligned}$$

as $\min(n_1, n_2) \to \infty$. This is only possible if for $j = 1, 2, ..., m$, $\lim_{\min(n_1,n_2)\to\infty} |c_{j,n_1} - c_{j,n_2}| = 0$. Thus, the $c_{j,n}$'s are Cauchy sequences in $\mathbb{R}$, and therefore converge to limits c_j. Denoting $y_m = \sum_{j=1}^m c_j x_j$, which is an element of $\mathcal{M}_m$, it follows now easily that $\lim_{n\to\infty} ||y_{n,m} - y_m|| = 0$. Thus, every Cauchy sequence in $\mathcal{M}_m$ converges to a limit in $\mathcal{M}_m$.

2.7.5 Theorem 2.4

Let X_n be a Cauchy sequence in $\mathcal{R}$. Then

$$||X_n - X_m||^2 = E\left[(X_n - X_m)^2\right] \to 0$$

as $\min(n, m) \to \infty$, so that by Chebyshev's inequality,

$$\underset{\min(n,m)\to\infty}{\text{plim}} |X_n - X_m| = 0.$$

As is well-known, convergence in probability is equivalent to almost sure (a.s.) convergence along a further subsequence of an arbitrary subsequence[4], hence there exists a subsequence n_k such that for $\min(k,m) \to \infty$,

$$|X_{n_k} - X_{n_m}| \overset{a.s.}{\to} 0.$$

In its turn this result is equivalent to the statement that there exists a set $N \in \mathcal{F}$ with $P(N) = 0$, called a null set, such that for all $\omega \in \Omega \backslash N$,

$$\lim_{\min(k,m)\to\infty} |X_{n_k}(\omega) - X_{n_m}(\omega)| = 0$$

Now $X_{n_k}(\omega)$ is a Cauchy sequence in $\mathbb{R}$ and thus converges to a limit $X(\omega)$ in $\mathbb{R}$, which is measurable $\mathcal{F}$,[5] so that X is a random variable defined on $\{\Omega, \mathcal{F}, P\}$. Hence, for fixed m and $k \to \infty$

$$(X_{n_k} - X_m)^2 \overset{a.s.}{\to} (X - X_m)^2 . \tag{2.10}$$

Finally, it follows from (2.10), Fatou's lemma[6] and the Cauchy property that

$$\begin{aligned} ||X - X_m||^2 &= E\left[(X - X_m)^2\right] = E\left[\lim_{k\to\infty} (X_{n_k} - X_m)^2\right] \\ &\leq \liminf_{k\to\infty} E\left[(X_{n_k} - X_m)^2\right] \to 0 \end{aligned}$$

for $m \to \infty$.

[4] See for example Bierens (2004, Theorem 6.B.3, p. 168).

[5] The latter follows from the well-known property that the limsup and liminf of a sequence of random variables are random variables themselves. See for example Bierens (2004, Theorem 2.13, p. 47).

[6] Fatou's lemma states: *For a sequence X_n of non-negative random variables,* $E[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} E[X_n]$. See for example Bierens (2004, Lemma 7.A.1, p. 201).

Chapter 3

Projections

3.1 The projection theorem

As is well-known from linear algebra and econometrics, the projection of a vector $y \in \mathbb{R}^n$ on the subspace spanned by vectors $x_1, ..., x_k$ in $\mathbb{R}^n$ is a linear combination $\widehat{y} = \sum_{j=1}^{k} \beta_j x_j$ such that $||y - \widehat{y}||$ is minimal. This is a linear regression problem: Minimize

$$||y - \widehat{y}||^2 = y'y - 2y'X\beta + \beta'X'X\beta$$

to $\beta = (\beta_1, ..., \beta_k)'$, where $X = (x_1, ..., x_k)$. If $k \leq n$ and the vectors $x_1, ..., x_k$ are linear independent then the solution is $\beta = (X'X)^{-1} X'y$, hence $\widehat{y} = X\beta = X(X'X)^{-1} X'y$.

If $x_1, ..., x_k$ are not linear independent then $\text{rank}(X) = m < k$. In that case we can rearrange $x_1, ..., x_k$ such that the matrix $X_1 = (x_1, ..., x_m)$ has rank m and $X_2 = (x_{m+1}, ..., x_k) = X_1 C$ for some $(k-m) \times (k-m)$ matrix C. Partition β accordingly as $\beta = (b_1', b_2')'$. Then

$$||y - \widehat{y}||^2 = y'y - 2y'X_1(b_1 - Cb_2) + (b_1 - Cb_2)' X_1'X_1(b_1 - Cb_2)$$

which is minimal for $(b_1 - Cb_2) = (X_1'X_1)^{-1} X_1'y$, hence

$$\widehat{y} = X_1 b_1 + X_2 b_2 = X_1(b_1 - Cb_2) = X_1(X_1'X_1)^{-1} X_1'y,$$

which is unique. The latter follows from Theorem 3.1 below.

The notion of a projection for Hilbert spaces is similar:

Definition 3.1. *The projection $\widehat{y}$ of an element y of a Hilbert space $\mathcal{H}$ on a subspace $\mathcal{S}$ is an element $\widehat{y} \in \mathcal{S}$ such that $||y - \widehat{y}|| = \inf_{z\in\mathcal{S}} ||y - z||$.*

However, we still have to show that $\widehat{y} \in \mathcal{S}$ is possible and unique. This follows from the fundamental projection theorem:

Theorem 3.1. (Projection theorem) *If $\mathcal{S}$ is a subspace of a Hilbert space $\mathcal{H}$ and y an element of $\mathcal{H}$ then there exists a unique element $\widehat{y} \in \mathcal{S}$ such that $||y - \widehat{y}|| = \inf_{z\in\mathcal{S}} ||y - z||$. Moreover the residual $u = y - \widehat{y}$ is orthogonal to any $z \in \mathcal{S}$: $\langle u, z\rangle = 0$.*

3.2 Projections in terms of angles

As is well known, the angle $\varphi(x, y)$ between two vectors x and y in a Euclidean space is defined by the cosine formula

$$\cos(\varphi(x, y)) = \frac{||x||^2 + ||y||^2 - ||x - y||^2}{2||x||.||y||} = \frac{x'y}{||x||.||y||},$$

due to the Law of Cosines.[1] Clearly, this formula carries over to elements x and y of a Hilbert space $\mathcal{H}$, simply by replacing the Euclidean inner product $x'y$ and norm $||x|| = \sqrt{x'x}$ by $\langle x, y\rangle$ and $||x|| = \sqrt{\langle x, x\rangle}$, respectively. Thus, the angle $\varphi(x, y)$ between two elements x and y of a Hilbert space is defined by the cosine formula

$$\cos(\varphi(x, y)) = \frac{\langle x, y\rangle}{||x||.||y||}. \tag{3.1}$$

Let $\mathcal{S}$, y and $\widehat{y}$ be as before, and let x be any element of $\mathcal{S}$. Then it follows from the cosine formula (3.1) and the orthogonality condition $\langle x, y - \widehat{y}\rangle = 0$ that

$$\cos(\varphi(x, y)) = \frac{\langle x, y\rangle}{||x||.||y||} = \frac{\langle x, \widehat{y}\rangle}{||x||.||y||} = \frac{||\widehat{y}||}{||y||} \cos(\varphi(x, \widehat{y}))$$

[1] Consider a triangle ABC, let φ be the angle between the legs $C \to A$ and $C \to B$, and denote the lengths of the legs opposite to the points A, B and C by α, β, and γ, respectively. Then $\gamma^2 = \alpha^2 + \beta^2 - 2\alpha\beta\cos(\varphi)$.

which is maximal if $\cos(\varphi(x,\widehat{y}))=1$. The latter is true if $x=c.\widehat{y}$ for some constant $c>0$. Consequently,

$$\cos(\varphi(y,\widehat{y}))=\max_{x\in\mathcal{S}}\cos(\varphi(x,y))=\frac{||\widehat{y}||}{||y||}. \tag{3.2}$$

3.3 Projections on subspaces spanned by a sequence

Let $\mathcal{H}$ be a Hilbert space and let $\{x_k\}_{k=1}^{\infty}$ be a sequence of elements of $\mathcal{H}$. Let $\mathcal{M}_n$ be the linear manifold spanned by $x_1,...,x_n$: $\mathcal{M}_n=\operatorname{span}(\{x_k\}_{k=1}^{n})$. As we have seen from Lemma 2.1, $\mathcal{M}_n$ is a Hilbert space.

Consider the projection $\widehat{y}_n$ of an element $y\in\mathcal{H}$ on $\mathcal{M}_n$. Then $\widehat{y}_n$ takes the form $\widehat{y}_n=\sum_{k=1}^{n}\theta_{n,k}x_k$, where the $\theta_{n,k}$'s are the solutions of the minimization problem

$$\begin{aligned}&\min_{\theta_1,\theta_2,...,\theta_n}\left\|y-\sum_{k=1}^{n}\theta_k x_k\right\|^2\\ &=\min_{\theta_1,\theta_2,...,\theta_n}\left(||y||^2-2\sum_{k=1}^{n}\theta_k\langle x_k,y\rangle+\sum_{k=1}^{n}\sum_{m=1}^{n}\theta_k\theta_m\langle x_k,x_m\rangle\right)\end{aligned}$$

Similar to linear regression, the first-order conditions involved are the normal equations

$$\sum_{m=1}^{n}\langle x_k,x_m\rangle\,\theta_{n,m}=\langle x_k,y\rangle\,,\ \ k=1,2,...,n,$$

which can be written in matrix-vector form as $\Sigma_{n,xx}\theta_n=\Sigma_{n,xy}$, for example. To solve this system uniquely as $\theta_n=\Sigma_{n,xx}^{-1}\Sigma_{n,xy}$ we need to impose a similar condition as linear independence in Euclidean spaces, namely **regularity**:

Definition 3.2. *Let $\{x_k\}_{k=1}^{\infty}$ be a sequence of elements of a Hilbert space $\mathcal{H}$. Denote the projection of x_k on* $\operatorname{span}\left(\{x_j\}_{j=k+1}^{\infty}\right)$ *by $\widehat{x}_k$, and let $u_k=x_k-\widehat{x}_k$. The sequence $\{x_k\}_{k=1}^{\infty}$ is said to be regular if $||u_k||>0$ for all $k\geq 1$.*

Exercise: Given a regular sequence $\{x_k\}_{k=1}^{\infty}$, prove that for $n=1,2,3,.....$ the $n\times n$ matrices $\Sigma_{n,xx}$ with elements $\langle x_i,x_j\rangle$ are nonsingular.

Lemma 3.1. *For* $z \in \text{span}(\{x_k\}_{k=1}^{\infty})$ *let* $\widehat{z}_n$ *be the projection of* z *on* $\text{span}(\{x_k\}_{k=1}^{n})$. *Then* $\lim_{n\to\infty} ||z - \widehat{z}_n|| = 0$.

More generally we have:

Theorem 3.2. *For* $z \in \mathcal{H}$, *let* $\widehat{z}$ *be the projection of* z *on* $\text{span}(\{x_k\}_{k=1}^{\infty})$ *and let* $\widehat{z}_n$ *be the projection of* z *on* $\text{span}(\{x_k\}_{k=1}^{n})$. *Then* $\lim_{n\to\infty} ||\widehat{z} - \widehat{z}_n|| = 0$.

Although each projection $\widehat{z}_n$ is a linear combination of $x_1, ..., x_n$, in general the result of Theorem 3.2 does **not** imply that there exists a sequence $\{\theta_j\}_{j=1}^{\infty}$ such that $\widehat{z} = \sum_{j=1}^{\infty} \theta_j x_j$. As an example of such a case, consider the Hilbert space $\mathcal{R}_0$ of zero-mean random variables with finite second moments, endowed with the inner product $\langle X, Y\rangle = E[X.Y]$ and associated norm and metric. Let

$$X_t = V_t - V_{t-1},$$

where V_t is distributed i.i.d. $N(0,1)$. This is clearly a zero-mean covariance stationary process, with covariance function $\gamma(0) = 2,\ \gamma(1) = -1,\ \gamma(m) = 0$ for $m \geq 2$. Hence $X_t \in \mathcal{R}_0$ for all t.

For given t, let $\mathcal{M}_{-\infty}^{t-1} = \text{span}(\{X_{t-m}\}_{m=1}^{\infty})$, $\mathcal{M}_{t-n}^{t-1} = \text{span}(X_{t-1}, ..., X_{t-n})$. The projection $\widehat{X}_{t,n}$ of X_t on $\mathcal{M}_{t-n}^{t-1}$ takes the form

$$\widehat{X}_{t,n} = \sum_{j=1}^{n} \theta_{n,j} X_{t-j}$$

where the coefficients $\theta_{n,j}$ are the solutions of the normal equations

$$\gamma(m) = \sum_{k=1}^{n} \gamma(|k-m|)\theta_{n,k},\ m = 1, ..., n.$$

hence for $n \geq 3$,

$$\begin{aligned}
-1 &= 2.\theta_{n,1} - \theta_{n,2} \\
0 &= -\theta_{n,1} + 2\theta_{n,2} - \theta_{n,3} \\
0 &= -\theta_{n,2} + 2\theta_{n,3} - \theta_{n,4} \\
&\vdots \\
0 &= -\theta_{n,n-2} + 2\theta_{n,n-1} - \theta_{n,n} \\
0 &= -\theta_{n,n-1} + 2\theta_{n,n}
\end{aligned}$$

The solutions of these normal equations are

$$\theta_{n,j} = \frac{j}{n+1} - 1,\ j = 1, ..., n,$$

hence

$$\widehat{X}_{t,n} = \sum_{j=1}^{n} \left(\frac{j}{n+1} - 1 \right) X_{t-j} \tag{3.3}$$

Next, let $\widehat{X}_t$ be the projection of X_t on $\mathcal{M}_{-\infty}^{t-1}$, and suppose that there exists a sequence $\{\theta_j\}_{j=1}^{\infty}$ such that $\widehat{X}_t = \sum_{j=1}^{\infty} \theta_j X_{t-j}$. Note that the latter is merely a short-hand notation for

$$\lim_{n\to\infty} \left\| \widehat{X}_t - \sum_{j=1}^{n} \theta_j X_{t-j} \right\|^2 = \lim_{n\to\infty} E\left[\left(\widehat{X}_t - \sum_{j=1}^{n} \theta_j X_{t-j} \right)^2 \right] = 0 \tag{3.4}$$

If so, it follows from Theorem 3.2 and (3.3) that

$$\begin{aligned} 0 &= \lim_{n\to\infty} E\left[\left(\sum_{j=1}^{n} \theta_j X_{t-j} - \sum_{j=1}^{n} \left(\frac{j}{n+1} - 1 \right) X_{t-j} \right)^2 \right] \\ &= \lim_{n\to\infty} E\left[\left(\sum_{j=1}^{n} \left(\frac{j}{n+1} - 1 - \theta_j \right) X_{t-j} \right)^2 \right] \end{aligned} \tag{3.5}$$

But

$$\begin{aligned} &\sum_{j=1}^{n} \left(\frac{j}{n+1} - 1 - \theta_j \right) X_{t-j} = \sum_{j=1}^{n} \left(\frac{j}{n+1} - 1 - \theta_j \right) (V_{t-j} - V_{t-j-1}) \\ &= -\left(\frac{n}{n+1} + \theta_1 \right) V_{t-1} - \sum_{j=1}^{n-1} \left(\theta_{j+1} - \theta_j - \frac{1}{n+1} \right) V_{t-j-1} \\ &+ \left(\frac{1}{n+1} + \theta_n \right) V_{t-n-1} \end{aligned}$$

hence

$$\begin{aligned} E\left[\left(\sum_{j=1}^{n} \left(\frac{j}{n+1} - 1 - \theta_j \right) X_{t-j} \right)^2 \right] &= \left(\frac{n}{n+1} + \theta_1 \right)^2 \\ &+ \sum_{j=1}^{n-1} \left(\theta_{j+1} - \theta_j - \frac{1}{n+1} \right)^2 + \left(\frac{1}{n+1} + \theta_n \right)^2 \end{aligned} \tag{3.6}$$

This equality implies that for arbitrary integers $m \geq 1$,

$$\begin{aligned}
&\liminf_{n\to\infty} E\left[\left(\sum_{j=1}^{n}\left(\frac{j}{n+1}-1-\theta_j\right)X_{t-j}\right)^2\right] \\
&\quad \geq \liminf_{n\to\infty}\left(\frac{n}{n+1}+\theta_1\right)^2+\liminf_{n\to\infty}\left(\theta_{m+1}-\theta_m-\frac{1}{n+1}\right)^2 \\
&\quad = (\theta_1+1)^2+(\theta_{m+1}-\theta_m)^2 .
\end{aligned}$$

Therefore, a necessary condition for (3.5) is that $\theta_m = -1$ for $m = 1, 2, 3, \ldots..$ But then it follows from (3.6) that

$$\lim_{n\to\infty} E\left[\left(\sum_{j=1}^{n}\left(\frac{j}{n+1}-1-\theta_j\right)X_{t-j}\right)^2\right] = \lim_{n\to\infty}\left(\frac{1}{n+1}-1\right)^2 = 1$$

which contradicts (3.5). Thus, in this case there does **not** exist a sequence $\{\theta_j\}_{j=1}^{\infty}$ such that (3.4) holds.

The problem that for the projection $\widehat{z}$ on $\text{span}\left(\{x_j\}_{j=1}^{\infty}\right)$ there does not always exist a sequence $\{\theta_j\}_{j=1}^{\infty}$ such that $\widehat{z} = \sum_{j=1}^{\infty}\theta_j x_j$ only occurs if the sequence $\{x_j\}_{j=1}^{\infty}$ is not orthogonal:

Theorem 3.3. *If a sequence $\{x_j\}_{j=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ is orthonormal, i.e.,*

$$\langle x_i, x_j\rangle = I(i=j), \tag{3.7}$$

then the projection $\widehat{z}$ of $z \in \mathcal{H}$ on $\text{span}(\{x_j\}_{j=1}^{\infty})$ *takes the form $\widehat{z} = \sum_{j=1}^{\infty}\theta_j x_j$ (in the sense that* $\lim_{n\to\infty}||\widehat{z}-\sum_{j=1}^{n}\theta_j x_j||$)*, where $\theta_j = \langle z, x_j\rangle$ with $\sum_{j=1}^{\infty}\theta_j^2 < \infty$.*

3.4 The Wold decomposition

Let $\mathcal{S}_1, \ldots, \mathcal{S}_n$ be subspaces of a Hilbert space $\mathcal{H}$. Then similar to Definition 2.10,

Definition 3.3. $\text{Span}(\mathcal{S}_1, \ldots, \mathcal{S}_n)$ *is the closure of the space of all linear combinations $\sum_{j=1}^{n} c_j x_j$, where $x_j \in \mathcal{S}_j$.*

We also need the definition of orthogonal complement:

Definition 3.4. *The orthogonal complement of a subspace $\mathcal{S}$ of a Hilbert space $\mathcal{H}$, denoted by $\mathcal{S}^{\perp}$, is the subset of $\mathcal{H}$ such that for each $x \in S$ and $y \in \mathcal{S}^{\perp}$, $\langle x, y \rangle = 0$.*

Lemma 3.2. *Orthogonal complements are subspaces.*

We can now formulate the following general version of the Wold decomposition:

Theorem 3.4. *Given a regular sequence $\{x_k\}_{k=1}^{\infty}$ in a Hilbert space, every $x \in \mathcal{S} = \text{span}(\{x_k\}_{k=1}^{\infty})$ can be written as $x = \sum_{k=1}^{\infty} \alpha_k e_k + w$, in the sense that $\lim_{n\to\infty} \|x - w - \sum_{k=1}^{n} \alpha_k e_k\| = 0$, where $\{e_k\}_{k=1}^{\infty}$ is an orthonormal sequence in $\mathcal{S}$, $\alpha_k = \langle x, e_k \rangle$, $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$, and*

$$w \in \mathcal{S}_{\infty} \cap \mathcal{U}_{\infty}^{\perp}, \tag{3.8}$$

with $\mathcal{S}_{\infty} = \cap_{n=1}^{\infty} \text{span}(\{x_k\}_{k=n}^{\infty})$ and $\mathcal{U}_{\infty}^{\perp}$ the orthogonal complement of $\mathcal{U}_{\infty} = \text{span}(\{e_k\}_{k=1}^{\infty})$. Note that (3.8) implies that w is orthogonal to all the e_k's: $\langle e_k, w \rangle = 0$ for $k = 1, 2, 3, ...$

In the case of the Hilbert space $\mathcal{R}_0$ of zero-mean random variables with finite second moments, with inner product $\langle X, Y \rangle = E[X.Y]$ and associated norm and metric, the results of Theorem 3.4 translate as follows:

Theorem 3.5. (*Wold decomposition theorem*) *Let X_t be a regular univariate zero-mean covariance stationary time series process. Then X_t can be written as*

$$X_t = \sum_{j=0}^{\infty} \alpha_j U_{t-j} + W_t \text{ a.s.}, \tag{3.9}$$

where U_t is a zero-mean uncorrelated process with variance 1,

$$\alpha_j = E[X_t U_{t-j}], \; \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \tag{3.10}$$

and W_t is a zero-mean covariance stationary process satisfying

$$W_t \in \mathcal{U}_t^{\perp} \cap \mathcal{S}_{-\infty}, \tag{3.11}$$

where $\mathcal{S}_{-\infty} = \cap_n \text{span}(\{X_{n-k}\}_{k=1}^{\infty})$ *and* $\mathcal{U}_t^{\perp}$ *is the orthogonal complement of* $\mathcal{U}_t = \text{span}(\{U_{t-k}\}_{k=0}^{\infty})$. *The result* (3.11) *implies that*

$$W_t \in \text{ span}\left(\{W_{t-m}\}_{m=1}^{\infty}\right), \tag{3.12}$$

which in its turn implies that W_t *is perfectly predictable from the past values* $W_{t-1}, W_{t-2}, W_{t-3}, \ldots\ldots.$ *Moreover,* (3.11) *implies that*

$$E[W_t U_{t-m}] = 0 \tag{3.13}$$

for all leads and lags m.

The condition $\text{var}(U_t) = 1$ is not essential as long as X_t is regular. Without loss of generality we may then replace U_t with $\widetilde{U}_t = \sigma U_t$, $\sigma > 0$, and α_k with $\widetilde{\alpha}_k/\sigma$, where σ can be pinned down by normalizing $\widetilde{\alpha}_0 = 1$.

The Wold decomposition carries over to k-variate covariance stationary processes X_t, as follows. Consider the Hilbert space $\mathcal{R}_k$ of zero mean random vectors in $\mathbb{R}^k$ with finite second moment matrices, endowed with the inner product $\langle X, Y\rangle = E\left[X'Y\right]$ and associated norm and metric. Let $\widehat{X}_t$ be the projection of X_t on $\text{span}(\{X_{t-j}\}_{j=1}^{\infty})$, with residual vector $V_t = X_t - \widehat{X}_t$, and let $\Sigma = E\left[V_t V_t'\right]$. In this case we need to extend the notion of regularity by requiring that Σ is positive definite rather than only $||V_t||^2 = E\left[V_t'V_t\right] > 0$, so that we can define $U_t = \Sigma^{-1/2}V_t$. Then the projection $\widetilde{X}_t$ of X_t on $\text{span}(\{U_{t-j}\}_{j=0}^{n})$ takes the form $\widetilde{X}_t = \sum_{j=1}^{n} A_j U_{t-j}$, where $A_j = E\left[X_t U_{t-j}'\right]$. It follows now straightforwardly from the proofs of Theorems 3.4 and 3.5 that

$$X_t = \sum_{j=1}^{\infty} A_j U_{t-j} + W_t \text{ a.s.},$$

where the process U_t is uncorrelated with zero expectation vector and variance matrix I_k, and $W_t \in \mathcal{U}_t^{\perp} \cap \mathcal{S}_{-\infty}$, with $\mathcal{U}_t^{\perp}$ and $\mathcal{S}_{-\infty}$ defined in Theorem 3.5.

It should be stressed that the deterministic process W_t is not necessarily nonrandom. For example let $W_t = a.\cos(\lambda t) + b.\sin(\lambda t)$, where a and b are independent random drawings from the standard normal distribution and $\lambda \in (-\pi/2, \pi/2)$ is a constant. Then $E\left[W_t\right] = 0$ and $E\left[W_t W_{t-m}\right] = \cos\left(\lambda m\right)$, hence W_t is a zero-mean covariance stationary process. If we observe W_{t-1}, W_{t-2} and W_{t-3} then we can solve a, b and λ, hence W_t is then determined for all t.

The question now arises under which conditions the deterministic process W_t is identical to zero. Since $W_t \in \cap_n \text{span}(\{X_{n-j}\}_{j=0}^{\infty})$, it follows that W_t is measurable with respect to the remote σ-algebra of the process X_t:

Definition 3.5. *Let $\mathcal{F}_t = \sigma(\{X_{t-j}\}_{j=0}^{\infty})$ be the σ-algebra generated by $\{X_{t-j}\}_{j=0}^{\infty}$. The σ-algebra $\mathcal{F}_{-\infty} = \cap_t \mathcal{F}_t$ is called the remote σ-algebra of the process X_t.*

If the process X_t is independent then it follows from Kolmogorov's zero-one law[2] that the sets in $\mathcal{F}_{-\infty}$ have either probability one or zero, so that the information in $\mathcal{F}_{-\infty}$ is non-informative. In other words, the memory of the remote past of X_t has vanished. However, this result carries over to certain dependent processes, for example α-mixing processes.[3] This gives rise to the notion of vanishing memory:

Definition 3.6. *A time series process is said to have a vanishing memory if the sets in its remote σ-algebra $\mathcal{F}_{-\infty}$ have either probability one or zero, i.e., $A \in \mathcal{F}_{-\infty}$ implies $P[A] = 1$ or $P[A] = 0$.*

In that case $E[W_t|\mathcal{F}_{-\infty}] = E[W_t]$ a.s.[4] However, since W_t is measurable $\mathcal{F}_{-\infty}$, we also have $E[W_t|\mathcal{F}_{-\infty}] = W_t$ a.s. Thus, $W_t = E[W_t] = 0$ a.s., where the second equality follows from the condition that $E[X_t] = 0$. Consequently,

Theorem 3.6. *If the zero-mean covariance stationary process X_t has a vanishing memory then the deterministic term W_t in its Wold decomposition is zero with probability 1.*

The Wold decomposition theorem in the form of Theorem 3.6 is the basis for time series analysis. In particular, for a univariate covariance stationary process X_t with a vanishing memory and expectation $E[X_t] = \mu$ the Wold decomposition can be written as

$$X_t = \mu + \alpha\,(L)\,U_t$$

where L is the lag operator and $\alpha\,(L) = 1 + \sum_{k=1}^{\infty} \alpha_k L^k$. The function $\alpha\,(L)$ can be approximated arbitrarily close by a ratio of two lag polynomials,

[2] See for example Bierens (2004, Theorem 7.5, p.185).
[3] See for example Bierens (2004, Theorem 7.6, p.186).
[4] See for example Bierens (2004, Exercise 3 in Section 7.6).

$\psi_q(L) = 1 + \sum_{k=1}^{q} \theta_k L^k$ and $\varphi_p(L) = 1 - \sum_{k=1}^{p} \gamma_k L^k$, of orders q and p, respectively, where at least $\varphi_p(L)$ is invertible with inverse $\varphi_p^{-1}(L)$.[5] In particular, for arbitrary $\varepsilon > 0$ there exist lag polynomials $\psi_q(L)$ and $\varphi_p(L)$ such that

$$E\left[\left(\left(\alpha(L) - \varphi_p^{-1}(L)\psi_q(L)\right)U_t\right)^2\right] < \varepsilon.$$

This gives rise to the well-known ARMA(p,q) models, for which it is assumed that $\alpha(L)$ is exactly of the form $\alpha(L) = \varphi_p^{-1}(L)\psi_q(L)$, so that $\varphi_p(L)X_t = \gamma + \psi_q(L)U_t$ with $\gamma_0 = \varphi_p(1)\mu$. Thus,

$$X_t = \gamma_0 + \sum_{k=1}^{p} \gamma_k X_{t-k} + U_t + \sum_{m=1}^{q} \theta_m U_{t-m}.$$

Moreover, if also $\psi_q(L)$ is invertible then X_t has the representation

$$\psi_q^{-1}(L)\varphi_p(L)X_t = \beta_0 + U_t,$$

where $\beta_0 = \mu.\varphi_p(1)/\psi_q(1)$. The lag function $\psi_q^{-1}(L)\varphi_p(L)$ can be written as $\psi_q^{-1}(L)\varphi_p(L) = 1 - \sum_{k=1}^{\infty} \beta_k L^k$, so that then X_t has the AR(∞) representation

$$X_t = \beta_0 + \sum_{k=1}^{\infty} \beta_k X_{t-k} + U_t.$$

This representation plays a key role in forecasting.

An important econometric application of the multivariate version of the Wold decomposition is Sims' (1980) innovation response analysis. Sims' (1980) landmark paper has changed the way empirical macroeconomics is conducted nowadays. His idea is the following. Let $X_t \in \mathbb{R}^k$ be a covariance stationary process of economic variables generated by a stationary VAR(p) process:

$$X_t = b_0 + \sum_{k=1}^{p} B_k X_{t-k} + U_t$$

Assume that the error vectors U_t are i.i.d. $N_k[0, \Sigma]$, where Σ is nonsingular. Stationarity of this process is equivalent to the requirement that the matrix-valued lag polynomial $B(L) = I_k - \sum_{k=1}^{p} B_k L^k$ is invertible.[6] The latter

[5] I.e., $\varphi_p(z) = 0$ for some $z \in \mathbb{C}$ implies $|z| > 1$.

[6] Which in its turn is equivalent to the condition that the roots of the polynomial $\det(B(z))$ are all outside the complex unit circle: $\det(B(z)) = 0$ implies $|z| > 1$.

condition also guarantees that X_t has a vanishing memory. It follows then from the Wold decomposition that X_t can be decomposed as

$$X_t = \mu + \sum_{m=0}^{\infty} A_m U_{t-m},$$

where $\mu = E[X_t]$ and $A_0 = I_k$. The parameters Σ, μ, and A_m can be estimated by estimating the VAR(p) for X_t by ordinary least squares, and then inverting the VAR(p) lag polynomial.

The variance matrix Σ of the U_t's can be written as $\Sigma = \Delta.\Delta'$, where Δ is a $k \times k$ lower-triangular matrix, so that U_t can be written as $U_t = \Delta e_t$, where now $e_t \sim N_k[0, I_k]$. Sims proposes to interpret the components of e_t as the unpredictable parts of policy interventions in the corresponding components of X_t. To trace the effect of these policy innovations on the future path of X_t, project X_{t+m} for $m \geq 0$ on component $e_{i,t}$ of e_t. These projections take the form $A_m \delta_i e_{i,t}$, where δ_i is column i of Δ, and may be interpreted as the response of X_{t+m} to the innovation $e_{i,t}$. Since the scale of $e_{i,t}$ does not matter, the responses of X_{t+m} for $m = 0, 1, 2, ...$ to a unit shock in $e_{i,t}$ are now $A_m \delta_i = E[X_{t+m} | e_{i,t} = 1] - E[X_{t+m}]$, which are usually presented in the form of graphs.

For more on the Wold decomposition and its time series applications, see for example Anderson (1994).

3.5 Projections on a random subspace

Because a Hilbert space $\mathcal{H}$ is a metric space, we can define open sets in $\mathcal{H}$ in the usual way. Therefore, similar to the Euclidean Borel field, we can define the Borel field $\mathcal{B}_{\mathcal{H}}$ of subsets of $\mathcal{H}$ as the smallest σ-algebra containing the collection of all open sets in $\mathcal{H}$, and call its elements Borel sets. Moreover, given a probability space $\{\Omega, \mathcal{F}, P\}$, where $\mathcal{F}$ is a σ-algebra of subsets of the sample space Ω and P is a probability measure on $\mathcal{F}$, a random element $X \in \mathcal{H}$ can now be defined as a mapping $X(.) : \Omega \to \mathcal{H}$ such that for all Borel sets $B \in \mathcal{B}_{\mathcal{H}}$, $\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$. In particular, if X_n is a sequence of random elements of $\mathcal{H}$ defined on a common probability space, the notion of convergence in probability of X_n to a non-random element x of $\mathcal{H}$, denoted by $\text{plim}_{n\to\infty} ||X_n - x||$, can be defined in the same way as for sequences of random vectors in a Euclidean space, i.e., for an arbitrary $\varepsilon > 0$,

$$\Pr\left[||X_n - x|| < \varepsilon\right] \stackrel{\text{def.}}{=} P\left(\{\omega \in \Omega : X(\omega) \in \{z \in \mathcal{H} : ||z - x|| < \varepsilon\}\}\right)$$
$$\rightarrow 1 \text{ as } n \rightarrow \infty,$$

because $\{z \in \mathcal{H} : ||z - x|| < \varepsilon\}$ is an open set and therefore a Borel set.

Now the question I will address is the following. Let Y_N and $X_{1,N}, ..., X_{n,N}$ be random elements of $\mathcal{H}$ depending on a sample of size N, where $n = n_N$ is a subsequence of N. Let $\widehat{Y}_{n,N}$ be the projection of Y_N on span$(X_{1,N}, ..., X_{n,N})$, and let $U_{n,N} = Y_N - \widehat{Y}_N$ be the residual. Under what conditions do $\widehat{Y}_{n,N}$ and $U_{n,N}$ converge in probability? The answer to this question is crucial for proving asymptotic normality of semi-nonparametric sieve estimators.

The answer is given in the following theorem.

Theorem 3.7. *Let Y_N and $X_{1,N}, X_{2,N}, ..., X_{n,N}$ be random elements of a Hilbert space $\mathcal{H}$ on the basis on a sample of size N, where n is a subsequence of N. Let $\widehat{Y}_{n,N}$ be the projection of Y_N on span$(\{X_{m,N}\}_{m=1}^{n})$, with residual $U_{n,N} = Y_N - \widehat{Y}_{n,N}$. Suppose that the following conditions hold.*
(a) There exists a non-random element y of $\mathcal{H}$ such that

$$\operatorname*{plim}_{N\rightarrow\infty} ||Y_N - y|| = 0. \tag{3.14}$$

(b) There exist a sequence $\{x_m\}_{m=1}^{\infty}$ of non-random elements of $\mathcal{H}$ and a sequence $\{\rho_m\}_{m=1}^{\infty}$ of positive numbers such that

$$\operatorname*{plim}_{N\rightarrow\infty} \sum_{m=1}^{n} \rho_m ||X_{m,N} - x_m|| = 0 \tag{3.15}$$

and

$$\liminf_{n\rightarrow\infty} \left\| \sum_{m=1}^{n} \rho_m x_m \right\| > 0. \tag{3.16}$$

Then $\text{plim}_{N\rightarrow\infty}||\widehat{Y}_{n,N} - \widehat{y}|| = 0$ *and* $\text{plim}_{N\rightarrow\infty}||U_{n,N} - u|| = 0$, *where $\widehat{y}$ is the projection of y on* span$(\{x_m\}_{m=1}^{\infty})$ *and $u = y - \widehat{y}$ is the residual involved.*

3.6 Appendix: Proofs

3.6.1 Theorem 3.1

Recall that "subspace" means a sub-Hilbert space. Thus, $\mathcal{S}$ is a Hilbert space.

Pick a sequence $z_n \in \mathcal{S}$ such that

$$||y - z_n|| \leq ||y - \widehat{y}|| + n^{-1}. \tag{3.17}$$

This is always possible because otherwise $||y - z|| > ||y - \widehat{y}|| + n^{-1}$ for all $z \in \mathcal{S}$ so that $\inf_{z \in \mathcal{S}} ||y - z|| \geq ||y - \widehat{y}|| + n^{-1}$. Then

$$\lim_{n\to\infty} ||y - z_n||^2 = ||y - \widehat{y}||^2 = \delta. \tag{3.18}$$

say. The first step is to show that z_n is a Cauchy sequence. Observe that

$$\begin{aligned} ||z_n - z_m||^2 &= ||(z_n - y) - (z_m - y)||^2 \\ &= ||z_n - y||^2 - 2\langle z_n - y, z_m - y\rangle + ||z_m - y||^2 \end{aligned}$$

and

$$\begin{aligned} 4.||0.5(z_n + z_m) - y||^2 &= ||(z_n - y) + (z_m - y)||^2 \\ &= ||z_n - y||^2 + 2\langle z_n - y, z_m - y\rangle + ||z_m - y||^2 \end{aligned}$$

Adding these two equation up yields

$$||z_n - z_m||^2 = 2||z_n - y||^2 + 2||z_m - y||^2 - 4.||0.5(z_n + z_m) - y||^2 \tag{3.19}$$

Because $0.5(z_n + z_m) \in \mathcal{S}$, it follows that $||0.5(z_n + z_m) - y||^2 \geq \delta^2$, whereas by (3.17) and (3.18), $||z_n - y||^2 \leq \left(\delta + n^{-1}\right)^2$ and $||z_m - y||^2 \leq \left(\delta + m^{-1}\right)^2$. Therefore, it follows from (3.19) that

$$\begin{aligned} ||z_n - z_m||^2 &\leq 2\left(\delta + n^{-1}\right)^2 + 2\left(\delta + m^{-1}\right)^2 - 4\delta^2 \\ &= 4\delta/n + 2n^{-2} + 4\delta/m + 2m^{-2}. \end{aligned}$$

Thus, z_n is a Cauchy sequence in $\mathcal{S}$ and therefore takes a limit $\widehat{y}$ in $\mathcal{S}$.

The next step is to show that for all $z \in \mathcal{S}$, $\langle y - \widehat{y}, z\rangle = 0$, as follows. Note that for any real scalar c, $\widehat{y} + c.z \in \mathcal{S}$ and therefore

$$||y - \widehat{y}||^2 \leq ||y - \widehat{y} - c.z||^2 = ||y - \widehat{y}||^2 - 2c.\langle y - \widehat{y}, z\rangle + c^2||z||^2$$

The right-hand side is minimal for $c = \langle y - \widehat{y}, z\rangle / ||z||^2$, hence

$$0 \leq -\frac{(\langle y - \widehat{y}, z\rangle)^2}{||z||^2}$$

and thus $\langle y - \widehat{y}, z\rangle = 0$.

Note that this argument only applies if the Hilbert space $\mathcal{H}$ is real. If $\mathcal{H}$ is complex this orthogonality proof can be adapted similar to the proof of Theorem 2.1.

Finally, we need to show that $\widehat{y}$ is unique. Suppose that there exists another projection $\widetilde{y} \in \mathcal{S}$. Then also $\langle y - \widetilde{y}, z\rangle = 0$, and thus $\langle y - \widetilde{y}, z\rangle - \langle y - \widehat{y}, z\rangle = \langle \widehat{y} - \widetilde{y}, z\rangle = 0$. But $z = y - \widetilde{y} \in \mathcal{S}$ so that $||\widehat{y} - \widetilde{y}||^2 = \langle \widehat{y} - \widetilde{y}, \widehat{y} - \widetilde{y}\rangle = 0$. Consequently, $\widehat{y}$ is unique.

3.6.2 Lemma 3.1

Let $\mathcal{M}_n = \text{span}(\{x_k\}_{k=1}^n)$ and $\mathcal{M}_\infty = \text{span}(\{x_k\}_{k=1}^\infty) = \overline{\cup_{n=1}^\infty \mathcal{M}_n}$. If $z \in \cup_{n=1}^\infty \mathcal{M}_n$ then there exists an n_0 such that $z \in \mathcal{M}_{n_0}$, hence for $n \geq n_0$, $\widehat{z}_n = z$ and thus $\lim_{n\to\infty} ||z - \widehat{z}_n|| = 0$. Now let $z \in \mathcal{M}_\infty \backslash (\cup_{n=1}^\infty \mathcal{M}_n)$. Since $\mathcal{M}_\infty = \overline{\cup_{n=1}^\infty \mathcal{M}_n}$ is closed and $\mathcal{M}_n \subset \mathcal{M}_{n+1}$, for each n there exists an $z_n \in \mathcal{M}_n$ such that $\lim_{n\to\infty} ||z - z_n||^2 = 0$, hence for $n \to \infty$, $||z - \widehat{z}_n||^2 \leq ||z - z_n||^2 \to 0$.

3.6.3 Theorem 3.2

Adopting the notation in the proof of Lemma 3.1, we may without loss of generality assume that $\widehat{z} \in \mathcal{M}_\infty \backslash (\cup_{n=1}^\infty \mathcal{M}_n)$, as otherwise the result of Theorem 3.2 holds trivially. Since $\mathcal{M}_\infty$ is closed this assumption implies that for each n we can select a $z_n \in \mathcal{M}_n$ such that

$$\lim_{n\to\infty} ||\widehat{z} - z_n|| = 0. \tag{3.20}$$

Let $||z - \widehat{z}|| = \delta$ and $||z - \widehat{z}_n|| = \delta_n$, and note that $\delta_n \geq \delta$. Since

$$\begin{aligned} \delta_n^2 &= ||z - \widehat{z}_n||^2 \leq ||z - z_n||^2 = ||z - \widehat{z} + \widehat{z} - z_n||^2 \\ &= ||z - \widehat{z}||^2 + ||\widehat{z} - z_n||^2 + 2\langle z - \widehat{z}, \widehat{z} - z_n\rangle \\ &= \delta^2 + ||\widehat{z} - z_n||^2 \end{aligned}$$

it follows from (3.20) that

$$\lim_{n\to\infty} \delta_n = \delta. \tag{3.21}$$

Recall that $z = \widehat{z} + u$, where $\langle u, x\rangle = 0$ for all $x \in \mathcal{M}_\infty$. Hence

$$\begin{aligned} ||\widehat{z} - \widehat{z}_n||^2 &= ||z - \widehat{z}_n - u||^2 = ||z - \widehat{z}_n||^2 + ||u||^2 - 2\langle z - \widehat{z}_n, u\rangle \\ &= ||z - \widehat{z}_n||^2 + ||u||^2 - 2\langle z, u\rangle = \delta_n^2 - \delta^2 \end{aligned} \tag{3.22}$$

where the last equality follows from $\langle z, u\rangle - \langle u, u\rangle = \langle \widehat{z}, u\rangle = 0$ and $\langle u, u\rangle = ||u||^2 = \delta^2$. The theorem now follows from (3.21) and (3.22).

3.6.4 Theorem 3.3

Due to the orthonormality condition (3.7), the projection $\widehat{z}_n$ of z on $\mathcal{M}_n = \text{span}(\{x_j\}_{j=1}^n)$ takes the form

$$\widehat{z}_n = \sum_{j=1}^n \theta_j x_j, \text{ where } \theta_j = \langle z, x_j\rangle. \tag{3.23}$$

Moreover, denoting $u_n = z - \widehat{z}_n$, it follows from (3.7) and (3.23) that

$$\begin{aligned} ||u_n||^2 &= \left\| z - \sum_{j=1}^n \theta_j x_j \right\|^2 = ||z||^2 - 2\sum_{j=1}^n \theta_j \langle z, x_j\rangle + \sum_{j=1}^n \sum_{i=1}^n \theta_j\theta_i \langle x_j, x_i\rangle \\ &= ||z||^2 - \sum_{j=1}^n \theta_j^2 \geq 0 \end{aligned} \tag{3.24}$$

hence $\sum_{j=1}^n \theta_j^2 \leq ||z||^2$ for all n and thus $\sum_{j=1}^\infty \theta_j^2 < \infty$. Finally, it follows from Theorem 3.2 that

$$\lim_{n\to\infty} \left\| \widehat{z} - \sum_{j=1}^n \theta_j x_j \right\|^2 = \lim_{n\to\infty} ||\widehat{z} - \widehat{z}_n||^2 = 0$$

so that we can write $\widehat{z} = \sum_{j=1}^\infty \theta_j x_j$.

3.6.5 Lemma 3.2

Let x be an arbitrary element of a subspace $\mathcal{S}$ of an Hilbert space $\mathcal{H}$ and let y_n be a Cauchy sequence in $\mathcal{S}^\perp$. Then there exists an $y \in \mathcal{H}$ such that $\lim_{n\to\infty} ||y - y_n|| = 0$. Since $\langle x, y_n\rangle = 0$ we have $\langle x, y\rangle = \langle x, y - y_n\rangle$. It follows now from the Cauchy-Schwarz inequality that $|\langle x, y\rangle| = |\langle x, y - y_n\rangle| \leq ||x||.||y - y_n|| \to 0$. Hence $y \in \mathcal{S}^\perp$.

3.6.6 Theorem 3.4

Denote $\mathcal{S}_n = \text{span}(\{x_k\}_{k=n}^{\infty})$. Project each x_k on $\mathcal{S}_{k+1}$, so that $x_k = \widehat{x}_k + u_k$ with projection $\widehat{x}_k \in \mathcal{S}_{k+1}$ and residual u_k. Recall that by the regularity condition, $||u_k|| > 0$, hence $e_k = u_k/||u_k||$ is well defined. It is not hard to verify that the residuals u_k are orthogonal, so that the e_k's are orthonormal. Next, denote

$$\mathcal{U}_n = \text{ span}\left(e_1, ..., e_n\right) = \text{ span}\left(u_1, ..., u_n\right),$$

and let $\mathcal{U}_n^{\perp}$ be the orthogonal complement of $\mathcal{U}_n$. Note that

$$\mathcal{U}_{n+1}^{\perp} \subset \mathcal{U}_n^{\perp}. \tag{3.25}$$

To see this, let $z \in \mathcal{U}_{n+1}^{\perp}$. Then for all $x \in \mathcal{U}_{n+1}$, $\langle z, x\rangle = 0$, and because obviously $\mathcal{U}_n \subset \mathcal{U}_{n+1}$, it follows that also $\langle z, x\rangle = 0$ for all $x \in \mathcal{U}_n$. Hence, $z \in \mathcal{U}_n^{\perp}$.

As before, let $\mathcal{M}_n = \text{span}(\{x_k\}_{k=1}^{n})$.

The theorem under review will be proved in six steps:

Step 1. First I will show that

$$\mathcal{M}_n \subset \text{ span}\left(\mathcal{U}_n, \mathcal{U}_n^{\perp} \cap \mathcal{S}_2\right). \tag{3.26}$$

Proof. Let $z \in \mathcal{M}_n$ be arbitrary. Recall that z takes the form $z = \sum_{k=1}^{n} c_k x_k$. Substituting $x_k = \widehat{x}_k + u_k = \widehat{x}_k + ||u_k||e_k$ we can write z as

$$\begin{aligned} z &= \sum_{k=1}^{n} c_k\left(\widehat{x}_k + u_k\right) = \sum_{k=1}^{n} c_k u_k + \sum_{k=1}^{n} c_k \widehat{x}_k \\ &= \sum_{k=1}^{n} c_k ||u_k|| e_k + \sum_{k=1}^{n} c_k \widehat{x}_k \end{aligned}$$

Note that

$$\sum_{k=1}^{n} c_k \widehat{x}_k \in \mathcal{S}_2 \tag{3.27}$$

because $\widehat{x}_k \in \mathcal{S}_{k+1} \subset \mathcal{S}_2$.

Next, project $\sum_{k=1}^{n} c_k \widehat{x}_k$ on $\mathcal{U}_n$. This projection takes the form $\widehat{p}_n = \sum_{k=1}^{n} d_k e_k$ with residual $w_{n+1} \in \mathcal{S}_2$. The latter follows from (3.27). But since

w_{n+1} is a residual of a projection on $\mathcal{U}_n$ we also have $\langle e_k, w_{n+1}\rangle = 0$ for $k = 1, ..., n$, hence $w_{n+1} \in \mathcal{U}_n^{\perp}$. Thus,

$$w_{n+1} \in \mathcal{U}_n^{\perp} \cap \mathcal{S}_2.$$

Denoting $\alpha_k = c_k||u_k|| + d_k$, we can now write

$$z = \sum_{k=1}^{n} \alpha_k e_k + w_{n+1}, \text{ where } w_{n+1} \in \mathcal{U}_n^{\perp} \cap \mathcal{S}_2.$$

Therefore, (3.26) holds.

Step 2. I will now show that

$$\text{span}\left(\mathcal{U}_n, \mathcal{U}_n^{\perp} \cap \mathcal{S}_2\right) = \text{span}\left(\mathcal{U}_n, \mathcal{U}_n^{\perp} \cap \mathcal{S}_{n+1}\right). \tag{3.28}$$

Proof. Denote

$$\mathcal{S}_{k,m} = \text{span}(\{x_j\}_{j=k}^{m})$$

for $m \geq k$ and let $z \in \mathcal{U}_n^{\perp} \cap \mathcal{S}_{2,m}$ for some $m \geq 2$. Consider first the case $m > n$. Since $z \in \mathcal{S}_{2,m}$ there exists constants c_k such that

$$\begin{aligned} z &= \sum_{k=2}^{m} c_k x_k = \sum_{k=2}^{n} c_k \left(\widehat{x}_k + u_k\right) + \sum_{k=n+1}^{m} c_k x_k \\ &= \sum_{k=2}^{n} c_k \left\|u_k\right\| e_k + \sum_{k=2}^{n} c_k \widehat{x}_k + \sum_{k=n+1}^{m} c_k x_k. \end{aligned}$$

Moreover, since $z \in \mathcal{U}_n^{\perp}$ it follows that $\langle z, e_k\rangle = 0$ for $k = 1, ..., n$. In particular,

$$\begin{aligned} 0 &= \langle z, e_2\rangle = c_2 \left\|u_2\right\| + \sum_{k=2}^{n} c_k \left\langle \widehat{x}_k, e_2\right\rangle + \sum_{k=n+1}^{m} c_k \left\langle x_k, e_2\right\rangle \\ &= c_2 \left\|u_2\right\| \end{aligned}$$

because $\sum_{k=2}^{n} c_k \widehat{x}_k \in S_3$, $\sum_{k=n+1}^{m} c_k x_k \in S_{n+1}$, and e_2 is orthogonal to S_3 and S_{n+1}. Hence $c_2 = 0$ and thus

$$z = \sum_{k=3}^{n} c_k \left\|u_k\right\| e_k + \sum_{k=3}^{n} c_k \widehat{x}_k + \sum_{k=n+1}^{m} c_k x_k.$$

It follows now similarly that $c_k = 0$ for $k = 3, ..., n$, hence

$$z = \sum_{k=n+1}^{m} c_k x_k \in \mathcal{S}_{n+1,m}.$$

Because $z \in \mathcal{U}_n^\perp$ as well, it follows now that

$$z \in \mathcal{U}_n^\perp \cap \mathcal{S}_{n+1,m},$$

which implies

$$\mathcal{U}_n^\perp \cap \mathcal{S}_{2,m} \subset \mathcal{U}_n^\perp \cap \mathcal{S}_{n+1,m}$$

because $z \in \mathcal{U}_n^\perp \cap \mathcal{S}_{2,m}$ was arbitrary. However, $\mathcal{S}_{n+1,m} \subset \mathcal{S}_{2,m}$ and therefore

$$\mathcal{U}_n^\perp \cap \mathcal{S}_{n+1,m} \subset \mathcal{U}_n^\perp \cap \mathcal{S}_{2,m},$$

so that

$$\mathcal{U}_n^\perp \cap \mathcal{S}_{2,m} = \mathcal{U}_n^\perp \cap \mathcal{S}_{n+1,m} \text{ for } m > n.$$

This result implies that

$$\mathcal{U}_n^\perp \cap \left(\cup_{m=n+1}^\infty \mathcal{S}_{2,m}\right) = \mathcal{U}_n^\perp \cap \left(\cup_{m=n+1}^\infty \mathcal{S}_{n+1,m}\right) \tag{3.29}$$

In the case $m \leq n$, $z \in \mathcal{U}_n^\perp \cap \mathcal{S}_{2,m}$ implies that $z = 0$, as can be straightforwardly verified from the above argument, so that $\mathcal{U}_n^\perp \cap \mathcal{S}_{2,m} = \{0\}$ for $m = 2, 3, ..., n$. Since Hilbert spaces are vector spaces and therefore always contain the null element it follows that

$$\begin{aligned} \cup_{m=2}^\infty \mathcal{U}_n^\perp \cap \mathcal{S}_{2,m} &= \{0\} \cup \left(\cup_{m=n+1}^\infty \mathcal{U}_n^\perp \cap \mathcal{S}_{2,m}\right) \\ &= \cup_{m=n+1}^\infty \mathcal{U}_n^\perp \cap \mathcal{S}_{2,m}, \end{aligned}$$

hence

$$\mathcal{U}_n^\perp \cap \left(\cup_{m=2}^\infty \mathcal{S}_{2,m}\right) = \mathcal{U}_n^\perp \cap \left(\cup_{m=n+1}^\infty \mathcal{S}_{2,m}\right). \tag{3.30}$$

Since by Definition 2.10,

$$\mathcal{S}_2 = \overline{\cup_{m=2}^\infty \mathcal{S}_{2,m}}, \ \mathcal{S}_{n+1} = \overline{\cup_{m=n+1}^\infty \mathcal{S}_{n+1,m}}$$

it follows now from (3.30) that

$$\begin{aligned} \mathcal{U}_n^\perp \cap \mathcal{S}_2 &= \mathcal{U}_n^\perp \cap \overline{\cup_{m=2}^\infty \mathcal{S}_{2,m}} \\ &= \mathcal{U}_n^\perp \cap \overline{\cup_{m=n+1}^\infty \mathcal{S}_{n+1,m}} = \mathcal{U}_n^\perp \cap \mathcal{S}_{n+1} \end{aligned}$$

which implies that (3.28) holds.

Step 3. Denote $\mathcal{R}_n = \text{span}\big(\mathcal{U}_n, \mathcal{U}_n^{\perp} \cap \mathcal{S}_{n+1}\big)$. Then

$$\mathcal{S}_1 = \overline{\cup_{n=1}^{\infty} \mathcal{R}_n}. \tag{3.31}$$

Proof. Combining (3.26) and (3.28) yields $\mathcal{M}_n \subset \mathcal{R}_n$, hence

$$\mathcal{S}_1 = \overline{\cup_{n=1}^{\infty} \mathcal{M}_n} \subset \overline{\cup_{n=1}^{\infty} \mathcal{R}_n}, \tag{3.32}$$

where the equality follows from Definition 2.10. However, we also have $\mathcal{R}_n \subset \mathcal{S}_1$, as is not hard to verify, hence

$$\overline{\cup_{n=1}^{\infty} \mathcal{R}_n} \subset \mathcal{S}_1. \tag{3.33}$$

Thus, the result (3.31) follows from (3.32) and (3.33).

Step 4. For an $x \in \mathcal{S}_1$, let $\widehat{x}_n$ be the projection of x on $\mathcal{R}_n$. Then

$$\widehat{x}_n = \sum_{j=1}^{n} \alpha_j e_j + w_{n+1} \tag{3.34}$$

where $\alpha_j = \langle x, e_j \rangle$ and w_{n+1} is the projection of x on $\mathcal{U}_n^{\perp} \cap S_{n+1}$. Moreover,

$$\sum_{j=1}^{\infty} \alpha_j^2 < \infty. \tag{3.35}$$

Furthermore,

$$\lim_{n \to \infty} \left\| x - \sum_{j=1}^{n} \alpha_j e_j - w_{n+1} \right\| = 0. \tag{3.36}$$

Proof. By the definition of $\mathcal{R}_n$ and by Definition 3.3, $\widehat{x}_n = \sum_{j=1}^{n} \theta_j e_j + w$ for some constants θ_j and a $w \in \mathcal{U}_n^{\perp} \cap S_{n+1}$. To determine the θ_j's and w, note that

$$\left\| x - \sum_{j=1}^{n} \theta_j e_j - w \right\|^2 = ||x - w||^2 - 2 \sum_{j=1}^{n} \theta_j \langle e_j, x \rangle + 2 \sum_{j=1}^{n} \theta_j \langle e_j, w \rangle$$

$$
\begin{aligned}
&+\left\|\sum_{j=1}^{n}\theta_j e_j\right\|^2 \\
&= \quad ||x-w||^2 - 2\sum_{j=1}^{n}\theta_j\left\langle e_j, x\right\rangle + \sum_{j=1}^{n}\theta_j^2
\end{aligned}
$$

because $w \in \mathcal{U}_n^{\perp} \cap S_{n+1} \subset \mathcal{U}_n^{\perp}$ implies $\langle e_j, w\rangle = 0$ and

$$
\left\|\sum_{j=1}^{n}\theta_j e_j\right\|^2 = \sum_{j=1}^{n}\sum_{i=1}^{n}\theta_j\theta_i\left\langle e_j, e_i\right\rangle = \sum_{j=1}^{n}\theta_j^2\left\langle e_j, e_j\right\rangle = \sum_{j=1}^{n}\theta_j^2.
$$

Thus

$$
\begin{aligned}
||x-\widehat{x}_n||^2 &= \inf_{\theta_1,\ldots,\theta_n, w\in\mathcal{U}_n^{\perp}\cap S_{n+1}}\left\|x-\sum_{j=1}^{n}\theta_j e_j - w\right\|^2 \\
&= \inf_{\theta_1,\ldots,\theta_n, w\in\mathcal{U}_n^{\perp}\cap S_{n+1}}\left(||x-w||^2 - 2\sum_{j=1}^{n}\theta_j\left\langle e_j, x\right\rangle + \sum_{j=1}^{n}\theta_j^2\right) \\
&= \inf_{w\in\mathcal{U}_n^{\perp}\cap S_{n+1}} ||x-w||^2 - \sum_{j=1}^{n}\alpha_j^2 \\
&= ||x-w_{n+1}||^2 - \sum_{j=1}^{n}\alpha_j^2 \qquad (3.37)
\end{aligned}
$$

where $\alpha_j = \langle x, e_j\rangle$ and w_{n+1} is the projection of x on $\mathcal{U}_n^{\perp} \cap S_{n+1}$.

This result implies that for all n,

$$
\sum_{j=1}^{n}\alpha_j^2 \leq ||x-w_{n+1}||^2 \leq ||x||^2 \qquad (3.38)
$$

so that (3.35) holds.

Finally, to prove (3.36), let $\widehat{x}$ be the projection of x on $\overline{\cup_{n=1}^{\infty}\mathcal{R}_n}$. Then it follows from Theorem 3.2 that $\lim_{n\to\infty}||\widehat{x}_n - \widehat{x}|| = 0$. But (3.31) implies $\widehat{x} \in \mathcal{S}_1$, hence $x = \widehat{x}$, so that $\lim_{n\to\infty}||\widehat{x}_n - x|| = 0$.

Step 5. Let $z_n = \sum_{j=1}^{n}\alpha_j e_j$. Then

$$
\lim_{n\to\infty}\|z-z_n\| = 0, \text{ where } z \in \mathcal{U}_\infty. \qquad (3.39)
$$

Proof. This follows from the fact that z_n is a Cauchy sequence in $\mathcal{U}_\infty = \text{span}(\{e_k\}_{k=1}^\infty)$ because

$$\begin{aligned} \|z_n - z_m\|^2 &= \left\| \sum_{j=\min(m,n)+1}^{\max(m,n)} \alpha_j e_j \right\|^2 \\ &= \sum_{j=\min(m,n)+1}^{\max(m,n)} \alpha_j^2 \le \sum_{j=\min(m,n)+1}^{\infty} \alpha_j^2 \\ &\to 0 \end{aligned}$$

as $\min(m,n) \to \infty$, where the latter is due to $\sum_{j=1}^\infty \alpha_j^2 < \infty$.

Step 6. There exists a $w \in \mathcal{U}_\infty^\perp \cap S_\infty$ such that

$$\lim_{n\to\infty} ||w_{n+1} - w|| = 0. \tag{3.40}$$

Proof. Recall from Step 4 that

$$w_{n+1} \in \mathcal{U}_n^\perp \cap S_{n+1}.$$

Moreover, it follows from (3.25) and the definition of S_{n+1} that for an arbitrary $k \ge 1$,

$$\mathcal{U}_n^\perp \cap S_{n+1} \subset \mathcal{U}_k^\perp \cap S_{k+1} \text{ for } n \ge k$$

hence

$$w_{n+1} \in \mathcal{U}_k^\perp \cap S_{k+1} \text{ for } n \ge k.$$

Furthermore for $n \ge k$, w_{n+1} is a Cauchy sequence in $\mathcal{U}_k^\perp \cap S_{k+1}$ because

$$\begin{aligned} ||w_{n+1} - w_{m+1}|| &= ||\widehat{x}_n - z_n - \widehat{x}_m + z_m|| \\ &\le ||\widehat{x}_n - \widehat{x}_m|| + ||z_n - z_m|| \\ &\le ||\widehat{x}_n - x|| + ||\widehat{x}_m - x|| + ||z_n - z_m|| \\ &\to 0 \end{aligned}$$

as $\min(m,n) \to \infty$. Thus, there exists a $w \in \mathcal{U}_k^\perp \cap S_{k+1}$ such that (3.40) holds. Since k was arbitrary we have $w \in \cap_{k=1}^\infty \mathcal{U}_k^\perp = \mathcal{U}_\infty^\perp$ and $w \in \cap_{k=1}^\infty S_{k+1} = S_\infty$, hence

$$w \in \mathcal{U}_\infty^\perp \cap S_\infty.$$

This completes the proof of Step 6.

The theorem now follows from (3.35), (3.39), (3.40) and the fact that $w \in \mathcal{U}_\infty^\perp \cap S_\infty \subset \mathcal{U}_\infty^\perp$, which implies that $\langle w, e_k \rangle = 0$ for $k \in \mathbb{N}$.

3.6.7 Theorem 3.5

Recall that $U_t = \widetilde{U}_t / \sqrt{E\left[\widetilde{U}_t^2\right]}$, where $\widetilde{U}_t = X_t - \widehat{X}_t$ with $\widehat{X}_t$ the projection of X_t on $\text{span}(\{X_{t-j}\}_{j=1}^\infty)$. The uncorrelatedness of the $\widetilde{U}_t$'s follows from Theorem 3.4, but we still need to show that $E[\widetilde{U}_t] = 0$ and $E[\widetilde{U}_t^2] = \sigma^2$ for all t.

Proof of $E[\widetilde{U}_t] = 0$
Let $\widehat{X}_{t,n}$ be the projection of X_t on $\text{span}(\{X_{t-j}\}_{j=1}^n)$. Then $\widehat{X}_{t,n}$ takes the form

$$\widehat{X}_{t,n} = \sum_{j=1}^n \beta_{j,n} X_{t-j},$$

where the $\beta_{j,n}$'s do not depend on t. The latter follows from the fact that the $\beta_{j,n}$'s are the solutions of the normal equations

$$\sum_{j=1}^n \beta_{j,n} \gamma(i-j) = \gamma(i), \;\; i = 1, 2, ..., n,$$

where $\gamma(i) = E\left[X_t X_{t-i}\right]$ is the covariance function of X_t. Hence $E\left[\widehat{X}_{t,n}\right] = 0$.

It follows from Theorem 3.2 that

$$\lim_{n\to\infty} \left\|\widehat{X}_{t,n} - \widehat{X}_t\right\|^2 = \lim_{n\to\infty} E\left[\left(\widehat{X}_{t,n} - \widehat{X}_t\right)^2\right] = 0 \tag{3.41}$$

so that by Liapounov's inequality and $E\left[\widehat{X}_{t,n}\right] = 0$,

$$\begin{aligned} \lim_{n\to\infty} \left|E[\widehat{X}_t]\right| &= \lim_{n\to\infty} \left|E[\widehat{X}_t - \widehat{X}_{t,n}]\right| \le \lim_{n\to\infty} E\left[\left|\widehat{X}_t - \widehat{X}_{t,n}\right|\right] \\ &\le \sqrt{\lim_{n\to\infty} E\left[\left(\widehat{X}_{t,n} - \widehat{X}_t\right)^2\right]} = 0. \end{aligned}$$

Thus $E[\widehat{X}_t] = 0$ and therefore $E[\widetilde{U}_t] = E[X_t - \widehat{X}_t] = 0$.

Proof of $E[\widetilde{U}_t^2] = \sigma^2$
Let $\widetilde{U}_{t,n} = X_t - \widehat{X}_{t,n}$. It follows from (3.41) that

$$\lim_{n\to\infty} E\left[\left(\widetilde{U}_t - \widetilde{U}_{t,n}\right)^2\right] = \lim_{n\to\infty} E\left[\left(\widehat{X}_{t,n} - \widehat{X}_t\right)^2\right] = 0 \tag{3.42}$$

Moreover,

$$\begin{aligned} E\left[\widetilde{U}_{t,n}^2\right] &= \left\|X_t - \widehat{X}_{t,n}\right\|^2 = E\left[\left(X_t - \sum_{j=1}^{n}\beta_{j,n}X_{t-j}\right)^2\right] \\ &= \gamma(0) - 2\sum_{j=1}^{n}\beta_{j,n}\gamma(j) + \sum_{j=1}^{n}\sum_{i=1}^{n}\beta_{j,n}\beta_{i,n}\gamma(i-j) \\ &= \sigma_n^2 \end{aligned}$$

say, which does not depend on t. Furthermore, note that σ_n^2 is non-increasing in n, so that

$$\lim_{n\to\infty}\sigma_n^2 = \sigma^2$$

exists, and that

$$\begin{aligned} E\left[\left(\widetilde{U}_t - \widetilde{U}_{t,n}\right)^2\right] &= \left\|\widehat{X}_{t,n} - \widehat{X}_t\right\|^2 = \left\|\widehat{X}_{t,n} - X_t + \widetilde{U}_t\right\|^2 \\ &= \left\|\widehat{X}_{t,n} - X_t\right\|^2 + 2\left\langle\widehat{X}_{t,n} - X_t, \widetilde{U}_t\right\rangle + ||\widetilde{U}_t||^2 \\ &= \left\|\widetilde{U}_{t,n}\right\|^2 - 2\left\langle X_t, \widetilde{U}_t\right\rangle + ||\widetilde{U}_t||^2 \\ &= \left\|\widetilde{U}_{t,n}\right\|^2 - 2\left\langle\widehat{X}_t + \widetilde{U}_t, \widetilde{U}_t\right\rangle + ||\widetilde{U}_t||^2 \\ &= \left\|\widetilde{U}_{t,n}\right\|^2 - 2\left\langle\widetilde{U}_t, \widetilde{U}_t\right\rangle + ||\widetilde{U}_t||^2 \\ &= \left\|\widetilde{U}_{t,n}\right\|^2 - ||\widetilde{U}_t||^2 \\ &= E\left[\widetilde{U}_{t,n}^2\right] - E\left[\widetilde{U}_t^2\right]. \end{aligned}$$

Thus,

$$E\left[\widetilde{U}_t^2\right] = \sigma_n^2 - E\left[\left(\widetilde{U}_t - \widetilde{U}_{t,n}\right)^2\right] \to \sigma^2.$$

Proof of (3.10), (3.11) and (3.13)
The result of Theorem 3.4 can now be translated as

$$\lim_{n\to\infty}\left\|X_t - \sum_{j=0}^{n}\alpha_j U_{t-j} - W_t\right\| = 0, \tag{3.43}$$

where U_t is a zero-mean uncorrelated covariance stationary process with unit variance, and $\alpha_k = \langle X_t, U_{t-k}\rangle = E\left[X_t U_{t-k}\right]$ with $\sum_{k=0}^{\infty}\alpha_k^2 < \infty$.

We still need to prove that the α_k's do not depend on t, as follows. Recall from the proof of $E[\widetilde{U}_t^2] = \sigma^2$ that $\widetilde{U}_{t,n} = X_t - \sum_{j=1}^{n}\beta_{j,n}X_{t-j}$, so that

$$E\left[X_{t+k}\widetilde{U}_{t,n}\right] = \gamma\left(k\right) - \sum_{j=1}^{n}\beta_{j,n}\gamma(k+j),$$

which does not depend on t. Moreover, by the Cauchy-Schwarz inequality and (3.42),

$$\lim_{n\to\infty}\left|E\left[X_{t+k}\left(\widetilde{U}_{t,n} - \widetilde{U}_t\right)\right]\right|^2 \leq \gamma\left(0\right)\lim_{n\to\infty}E\left[\left(\widetilde{U}_{t,n} - \widetilde{U}_t\right)^2\right] = 0.$$

Thus $E\left[X_{t+k}\widetilde{U}_t\right] = \lim_{n\to\infty}E\left[X_{t+k}\widetilde{U}_{t,n}\right]$. Since the latter does not depend on t, neither does $\alpha_k = E\left[X_{t+k}U_t\right] = E\left[X_{t+k}\widetilde{U}_t/||\widetilde{U}_t||\right]$.

The results (3.11) and (3.13) follow straightforwardly from Theorem 3.4.

Proof of (3.9)
The result (3.43) implies, by Chebyshev's inequality, that

$$X_t = \operatorname*{plim}_{n\to\infty}\sum_{j=0}^{n}\alpha_j U_{t-j} + W_t. \tag{3.44}$$

Recall that convergence in probability for $n \to \infty$ is equivalent to a.s. convergence along a further subsequence k_m of an arbitrary subsequence of n. See for example Bierens (2004, Theorem 6.B.3, p. 168). Thus for such a subsequence k_m,

$$\sum_{j=0}^{k_m}\alpha_j U_{t-j} \stackrel{a.s.}{\to} X_t - W_t \tag{3.45}$$

as $m \to \infty$, and the same holds for any further subsequence of k_m.

Without loss of generality we may choose $k_0 = 0$. Then for each $n > 0$ we can find an m_n such that

$$k_{m_{n-1}} < n \leq k_{m_n}. \tag{3.46}$$

Moreover, (3.45) implies that

$$\sum_{j=0}^{k_{m_n}} \alpha_j U_{t-j} \stackrel{a.s.}{\to} X_t - W_t \text{ as } n \to \infty. \tag{3.47}$$

Due to (3.46),

$$\sum_{n=1}^{\infty} E\left[\left(\sum_{j=0}^{k_{m_n}} \alpha_j U_{t-j} - \sum_{j=0}^{n} \alpha_j U_{t-j}\right)^2\right] = \sum_{n=1}^{\infty} E\left[\left(\sum_{j=n+1}^{k_{m_n}} \alpha_j U_{t-j}\right)^2\right]$$
$$\leq \sum_{n=1}^{\infty} \sum_{j=k_{m_{n-1}}+1}^{k_{m_n}} \alpha_j^2 \leq \sum_{j=0}^{\infty} \alpha_j^2 < \infty,$$

so that by Chebyshev's inequality, for arbitrary $\varepsilon > 0$,

$$\sum_{n=0}^{\infty} \Pr\left[\left|\sum_{j=0}^{k_{m_n}} \alpha_j U_{t-j} - \sum_{j=0}^{n} \alpha_j U_{t-j}\right| > \varepsilon\right] < \infty.$$

This result implies, by the Borel-Cantelli lemma,[7] that

$$\sum_{j=0}^{k_{m_n}} \alpha_j U_{t-j} - \sum_{j=0}^{n} \alpha_j U_{t-j} \stackrel{a.s.}{\to} 0 \text{ as } n \to \infty. \tag{3.48}$$

Combining (3.47) and (3.48) it follow now that

$$\sum_{j=0}^{n} \alpha_j U_{t-j} \stackrel{a.s.}{\to} X_t - W_t \text{ as } n \to \infty. \tag{3.49}$$

Since $\sum_{j=0}^{\infty} \alpha_j U_{t-j}$ is defined as $\lim_{n\to\infty} \sum_{j=0}^{n} \alpha_j U_{t-j}$, (3.9) is equivalent to (3.49).

[7] See for example Bierens (2004, Theorem 6.B.2, p. 168).

The zero-mean covariance stationarity of W_t
It follows now trivially from (3.9) that $E[W_t] = 0$. Moreover, it is left as an exercise to show that for $m \geq 0$,

$$E[W_t W_{t-m}] = \gamma(m) - \sum_{j=0}^{\infty} \alpha_{j+m}\alpha_j. \tag{3.50}$$

Proof of (3.12)
Finally, $W_t \in \cap_n \text{span}(\{X_{n-j}\}_{j=0}^{\infty})$ implies that $W_t \in \text{span}(\{X_{t-j}\}_{j=1}^{\infty})$, hence the projection of W_t on $\text{span}(\{X_{t-j}\}_{j=1}^{\infty})$ is W_t itself. Since by (3.9),

$$\text{span}(\{X_{t-j}\}_{j=1}^{\infty}) = \text{span}\left(\text{span}(\{U_{t-j}\}_{j=1}^{\infty}), \text{span}(\{W_{t-j}\}_{j=1}^{\infty})\right)$$

and the projection of W_t on $\text{span}(\{U_{t-j}\}_{j=1}^{\infty})$ is zero, it follows that the projection of W_t on $\text{span}(\{W_{t-j}\}_{j=1}^{\infty})$ is W_t itself, which proves (3.12). ■

3.6.8 Theorem 3.7

Note that $||\widehat{Y}_{n,N} - \widehat{y}|| = ||(u - U_{n,N}) - (Y_N - y)|| \leq ||U_{n,N} - u|| + ||Y_N - y||$, hence

$$||\widehat{Y}_{n,N} - \widehat{y}|| \leq ||U_{n,N} - u|| + o_p(1),$$

where the $o_p(1)$ term follows from condition (a). Therefore it suffices to prove $||U_{n,N} - u|| = o_p(1)$, as follows.

Let $\widetilde{Y}_{n,N}$ be the projection of y on $\text{span}(\{X_{m,N}\}_{m=1}^{n})$, with residual $\widetilde{U}_{n,N} = y - \widetilde{Y}_{n,N}$, and let $\widehat{y}_n$ be the projection of y on $\text{span}(\{x_m\}_{m=1}^{n})$, with residual $u_n = y - \widehat{y}_n$. Then by the triangular inequality,

$$||U_{n,N} - u_n|| \leq ||U_{n,N} - \widetilde{U}_{n,N}|| + ||u_n - \widetilde{U}_{n,N}||.$$

It will be shown that

$$||U_{n,N} - \widetilde{U}_{n,N}|| = o_p(1) \tag{3.51}$$

and

$$||\widehat{y}_n - \widetilde{Y}_{n,N}|| = ||u_n - \widetilde{U}_{n,N}|| = o_p(1). \tag{3.52}$$

Since $\lim_{n\to\infty} ||\widehat{y}_n - \widehat{y}|| = 0$ and thus $\lim_{n\to\infty} ||u_n - u|| = 0$, the result of the theorem under review then follows from (3.51) and (3.52).

Proof of (3.51)

Denote the angle between two elements x and y of $\mathcal{H}$ by $\varphi(x, y)$. Recall that

$$\begin{aligned}
\sin^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) &= ||U_{n,N}||^2/||Y_N||^2 \\
\sin^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) &= ||\widetilde{U}_{n,N}||^2/||y||^2 \\
\cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) &= ||\widehat{Y}_{n,N}||/||Y_N|| \\
\cos\left(\varphi(y, \widetilde{Y}_{n,N})\right) &= ||\widetilde{Y}_{n,N}||/||y||.
\end{aligned}$$

Using these formulas we can write

$$\begin{aligned}
||U_{n,N} - \widetilde{U}_{n,N}||^2 &= ||U_{n,N}||^2 + ||\widetilde{U}_{n,N}||^2 - 2\left\langle U_{n,N}, \widetilde{U}_{n,N}\right\rangle \\
&= ||Y_N||^2 \sin^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) + ||y||^2 \sin^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) \\
&\quad -2\left\langle U_{n,N}, \widetilde{U}_{n,N}\right\rangle \\
&= ||Y_N||^2 + ||y||^2 \\
&\quad -||Y_N||^2 \cos^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) - ||y||^2 \cos^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) \\
&\quad -2\left\langle U_{n,N}, \widetilde{U}_{n,N}\right\rangle \\
&= ||Y_N - y||^2 - ||Y_N||^2 \cos^2\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) \\
&\quad -||y||^2 \cos^2\left(\varphi(y, \widetilde{Y}_{n,N})\right) + 2\left\langle Y_N, y\right\rangle - 2\left\langle U_{n,N}, \widetilde{U}_{n,N}\right\rangle
\end{aligned}$$

and

$$\begin{aligned}
\left\langle U_{n,N}, \widetilde{U}_{n,N}\right\rangle &= \langle U_{n,N}, \widetilde{U}_{n,N} + \widetilde{Y}_{n,N}\rangle = \langle U_{n,N}, y\rangle \\
&= \langle U_{n,N} + \widehat{Y}_{n,N}, y\rangle - \langle \widehat{Y}_{n,N}, y\rangle \\
&= \langle Y_N, y\rangle - \langle \widehat{Y}_{n,N}, y\rangle \\
&= \langle Y_N, y\rangle - \cos\left(\varphi(y, \widehat{Y}_{n,N})\right) ||\widehat{Y}_{n,N}||.||y|| \\
&= \langle Y_N, y\rangle - \cos\left(\varphi(y, \widehat{Y}_{n,N})\right) \cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) ||y||.||Y_N|| \\
&\geq \langle Y_N, y\rangle - \cos\left(\varphi(y, \widetilde{Y}_{n,N})\right) \cos\left(\varphi(Y_N, \widehat{Y}_{n,N})\right) ||y||.||Y_N||
\end{aligned}$$

where the inequality follows from $\cos\left(\varphi(y,\widehat{Y}_{n,N})\right) \le \cos\left(\varphi(y,\widetilde{Y}_{n,N})\right)$. Thus

$$\begin{aligned} ||U_{n,N}-\widetilde{U}_{n,N}||^2 &\le ||Y_N-y||^2 \\ &\quad -\left(||Y_N||\cos\left(\varphi(Y_N,\widehat{Y}_{n,N})\right)-||y||\cos\left(\varphi(y,\widetilde{Y}_{n,N})\right)\right)^2 \\ &\le ||Y_N-y||^2 = o_p(1) \end{aligned}$$

where the $o_p(1)$ term is due to condition (3.14). This proves (3.51).

Proof of (3.52)
Let $r_1 = x_1$ and for $m \ge 2$, let r_m be the residual of the projection of x_m on span$(x_1,...,x_{m-1})$. Denote $e_m = ||r_m||^{-1}r_m$ if $||r_m|| > 0$ and $e_m = 0$ if $||r_m|| = 0$. Similarly, let $R_{1,N} = X_{1,N}$ and for $m = 2,...,n$, let $R_{m,N}$ be the residual of the projection of $X_{m,N}$ on span$(X_{1,N},...,X_{m-1,N})$. Denote $\widehat{e}_{m,N} = ||R_{m,N}||^{-1}R_{m,N}$ if $||R_{m,N}|| > 0$, and $\widehat{e}_{m,N} = 0$ if $||R_{m,N}|| = 0$. Then we can write

$$\widehat{y}_n = \sum_{m=1}^{n}\alpha_m e_m, \text{ where } \alpha_m = \langle y, e_m\rangle \text{ and } \sum_{m=1}^{\infty}\alpha_m^2 < \infty \quad (3.53)$$

$$\widetilde{Y}_{n,N} = \sum_{m=1}^{n}\widehat{\alpha}_{m,N}\widehat{e}_{m,N}, \text{ where } \widehat{\alpha}_{m,N} = \langle y, \widehat{e}_{m,N}\rangle. \quad (3.54)$$

It follows from the trivial equalities $||\widehat{y}_n - \widetilde{Y}_{n,N}||^2 = ||\widetilde{Y}_{n,N}||^2 + ||\widehat{y}_n||^2 - 2\left\langle \widehat{y}_n, \widetilde{Y}_{n,N}\right\rangle$ and $\left\langle \widehat{y}_n, \widetilde{Y}_{n,N}\right\rangle = \left\langle \widehat{y}_n, y-\widetilde{U}_{n,N}\right\rangle = ||\widehat{y}_n||^2 - \left\langle \widehat{y}_n, \widetilde{U}_{n,N}\right\rangle$ that

$$||\widehat{y}_n - \widetilde{Y}_{n,N}||^2 = ||\widetilde{Y}_{n,N}||^2 - ||\widehat{y}_n||^2 + 2\left\langle \widehat{y}_n, \widetilde{U}_{n,N}\right\rangle.$$

Moreover, using the Cauchy-Schwarz inequality and the fact that $||\widetilde{U}_{n,N}|| \le ||y||$, it follows that

$$\begin{aligned} \left|\left\langle \widehat{y}_n, \widetilde{U}_{n,N}\right\rangle\right| &= \left|\left\langle \sum_{m=1}^{n}\alpha_m e_m, \widetilde{U}_{n,N}\right\rangle\right| = \left|\left\langle \sum_{m=1}^{n}\alpha_m (e_m - \widehat{e}_{m,N}), \widetilde{U}_{n,N}\right\rangle\right| \\ &\le ||\widetilde{U}_{n,N}||.\left\|\sum_{m=1}^{n}\alpha_m(e_m-\widehat{e}_{m,N})\right\| \\ &\le ||y||.\left\|\sum_{m=1}^{n}\alpha_m(e_m-\widehat{e}_{m,N})\right\| \end{aligned}$$

$$\leq \quad ||y||.\left\|\sum_{m=1}^{k}\alpha_m(e_m-\widehat{e}_{m,N})\right\|+2||y||\sqrt{\sum_{m=k+1}^{\infty}\alpha_m^2}$$

Given an arbitrary $\varepsilon > 0$ we can choose k so large that $2||y||\sqrt{\sum_{m=k+1}^{\infty}\alpha_m^2} < \varepsilon$, and for this k, $\left\|\sum_{m=1}^{k}\alpha_m(e_m-\widehat{e}_{m,N})\right\| = o_p(1)$, as is easy to verify from condition (3.15). Consequently,

$$\left\langle \widehat{y}_n, \widetilde{U}_{n,N}\right\rangle = o_p(1)$$

and thus

$$||\widehat{y}_n - \widetilde{Y}_{n,N}||^2 = ||\widetilde{Y}_{n,N}||^2 - ||\widehat{y}_n||^2 + o_p(1). \tag{3.55}$$

The next step is to show that

$$||\widetilde{Y}_{n,N}|| \leq ||\widehat{y}_n|| + o_p(1), \tag{3.56}$$

as follows. Note that

$$\begin{aligned}
||\widetilde{U}_{n,N}|| &= \inf_{\beta_1,\ldots,\beta_n}\left\|y-\sum_{m=1}^{n}\beta_m X_{m,N}\right\|^2 \\
&= \inf_{(\xi_1,\ldots,\xi_n)'\in\times_{m=1}^{n}[-\rho_m,\rho_m]}\inf_{\lambda}\left\|y-\lambda\sum_{m=1}^{n}\xi_m X_{m,N}\right\|^2 \\
&= \inf_{(\xi_1,\ldots,\xi_n)'\in\times_{m=1}^{n}[-\rho_m,\rho_m]}\left\{||y||^2-\left(\frac{\langle y,\sum_{m=1}^{n}\xi_m X_{m,N}\rangle}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|}\right)^2\right\} \\
&= ||y||^2 - \sup_{(\xi_1,\ldots,\xi_n)'\in\times_{m=1}^{n}[-\rho_m,\rho_m]}\frac{\left(\langle y,\sum_{m=1}^{n}\xi_m X_{m,N}\rangle\right)^2}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|^2}
\end{aligned}$$

hence

$$||\widetilde{Y}_{n,N}|| = \sup_{(\xi_1,\ldots,\xi_n)'\in\times_{m=1}^{n}[-\rho_m,\rho_m]}\frac{\langle y,\sum_{m=1}^{n}\xi_m X_{m,N}\rangle}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|} \tag{3.57}$$

and similarly,

$$||\widehat{y}_n|| = \sup_{(\xi_1,\ldots,\xi_n)'\in\times_{m=1}^{n}[-\rho_m,\rho_m]}\frac{\langle y,\sum_{m=1}^{n}\xi_m x_m\rangle}{\|\sum_{m=1}^{n}\xi_m x_m\|} \tag{3.58}$$

Note that by condition (3.16) at least one x_m is non-zero, so that (3.57) and (3.58) are well-defined for sufficiently large n.

Since the ratios in (3.57) and (3.58) are scale-invariant, we may without loss of generality impose the normalization

$$\left\|\sum_{m=1}^{n}\xi_m x_m\right\| = M_n = \frac{1}{2}\left\|\sum_{m=1}^{n}\rho_m x_m\right\|, \tag{3.59}$$

for example. Note that (3.59) is compatible with $(\xi_1,...,\xi_n)' \in \mathsf{X}_{m=1}^{n}[-\rho_m,\rho_m]$. Thus, denoting

$$\Xi_n = \left\{(\xi_1,...,\xi_n)' \in \mathsf{X}_{m=1}^{n}[-\rho_m,\rho_m] : \left\|\sum_{m=1}^{n}\xi_m x_m\right\| = M_n\right\},$$

the expressions (3.57) and (3.58) are equivalent to

$$||\widetilde{Y}_{n,N}|| = \sup_{(\xi_1,...,\xi_n)'\in\Xi_n} \frac{\langle y, \sum_{m=1}^{n}\xi_m X_{m,N}\rangle}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|} \tag{3.60}$$

and

$$||\widehat{y}_n|| = \sup_{(\xi_1,...,\xi_n)'\in\Xi_n} \frac{\langle y, \sum_{m=1}^{n}\xi_m x_m\rangle}{\|\sum_{m=1}^{n}\xi_m x_m\|}, \tag{3.61}$$

respectively.

Finally, observe from (3.59), (3.60) and (3.61) that

$$\begin{aligned}
||\widehat{y}_n|| &= \sup_{(\xi_1,...,\xi_n)'\in\Xi_n}\left\{\frac{\langle y, \sum_{m=1}^{n}\xi_m X_{m,N}\rangle}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|} \times \frac{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|}{\|\sum_{m=1}^{n}\xi_m x_m\|}\right. \\
&\qquad\qquad \left. - \frac{\langle y, \sum_{m=1}^{n}\xi_m (X_{m,N}-x_m)\rangle}{\|\sum_{m=1}^{n}\xi_m x_m\|}\right\} \\
&= \sup_{(\xi_1,...,\xi_n)'\in\Xi_n}\left\{\frac{\langle y, \sum_{m=1}^{n}\xi_m X_{m,N}\rangle}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|} \times \frac{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|}{M_n}\right. \\
&\qquad\qquad \left. - \frac{\langle y, \sum_{m=1}^{n}\xi_m (X_{m,N}-x_m)\rangle}{M_n}\right\} \\
&\geq \left(1 - \frac{\sum_{m=1}^{n}\rho_m ||X_{m,N}-x_m||}{M_n}\right) \sup_{(\xi_1,...,\xi_n)'\in\Xi_n} \frac{\langle y, \sum_{m=1}^{n}\xi_m X_{m,N}\rangle}{\|\sum_{m=1}^{n}\xi_m X_{m,N}\|} \\
&\quad -||y||\frac{\sum_{m=1}^{n}\rho_m ||X_{m,N}-x_m||}{M_n}
\end{aligned}$$

$$
\begin{aligned}
&= \left(1 - \frac{\sum_{m=1}^{n} \rho_m ||X_{m,N} - x_m||}{M_n}\right) ||\widetilde{Y}_{n,N}|| - ||y|| \frac{\sum_{m=1}^{n} \rho_m ||X_{m,N} - x_m||}{M_n} \\
&\geq ||\widetilde{Y}_{n,N}|| - 2||y|| \frac{\sum_{m=1}^{n} \rho_m ||X_{m,N} - x_m||}{M_n}, \qquad (3.62)
\end{aligned}
$$

where the last inequality follows from $||\widetilde{Y}_{n,N}|| \leq ||y||$. Since by condition (3.16), $\liminf_{n\to\infty} M_n > 0$, it follows from (3.15) and (3.62) that (3.56) holds. The latter together with (3.55) imply (3.52).

Chapter 4

Orthogonal polynomials

4.1 Introduction

Let $w(x)$ be a non-negative Borel measurable real-valued function on $\mathbb{R}$ satisfying

$$\int_{-\infty}^{\infty} |x|^k \, w(x)dx \in (0, \infty) \text{ for } k \in \mathbb{N}_0$$

where the integral involved is the Lebesgue integral. Without loss of generality we may assume that w is a density function with finite absolute moments of any order. Let

$$p_k(x|w) = \sum_{j=0}^{k} \alpha_{k,j} x^j, \ \alpha_{k,k} = 1, \ k \in \mathbb{N}_0 \tag{4.1}$$

be a sequence of polynomials in $x \in \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} p_k(x|w) p_m(x|w) w(x) dx = 0 \text{ if } k \neq m. \tag{4.2}$$

In words, the polynomials $p_k(x|w)$ are *orthogonal* with respect to the weight function $w(x)$.

Defining

$$\overline{p}_k(x|w) = \frac{p_k(x|w)}{\sqrt{\int_{-\infty}^{\infty} p_k(y|w)^2 w(y) dy}} \tag{4.3}$$

yields a sequence of *orthonormal* polynomials w.r.t. $w(x)$:

$$\int_{-\infty}^{\infty} \overline{p}_k(x|w)\overline{p}_m(x|w)w(x)dx = \begin{cases} 0 & \text{if } k \neq m, \\ 1 & \text{if } k = m. \end{cases} \tag{4.4}$$

This sequence is 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 $j < k$,

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

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.

The reason for considering orthonormal polynomials is the following.

Theorem 4.1. *Let $w(x)$ be a density function with support (a, b), $-\infty \leq a < b \leq \infty$, satisfying the moment conditions*

$$\int_{-\infty}^{\infty} |x|^k\, w(x)dx < \infty \tag{4.5}$$

for $k \in \mathbb{N}_0$. Denote by $L^2(w)$ be the Hilbert space of Borel measurable real functions f on (a,b) satisfying $\int_a^b f(x)^2 w(x)dx < \infty$, with inner product $\langle f,g\rangle = \int_a^b f(x)g(x)w(x)dx$ and associated norm $\|f\| = \sqrt{\langle f,f\rangle}$ and metric $||f-g||$. For an arbitrary function $f \in L^2(w)$, let

$$f_n(x) = \sum_{k=0}^{n} \gamma_k \overline{p}_k(x|w),$$

where

$$\gamma_k = \langle f, \overline{p}_k\rangle = \int_a^b f(x)\overline{p}_k(x|w)w(x)dx. \tag{4.6}$$

Then

$$\lim_{n\to\infty} \|f - f_n\| = 0. \tag{4.7}$$

This result implies that every function $f \in L^2(w)$ can be written as

$$f(x) = \sum_{k=0}^{\infty} \gamma_k \overline{p}_k(x|w) \ \text{a.e. on } (a,b). \tag{4.8}$$

Note that condition (4.5) holds trivially if the support (a,b) of $w(x)$ is bounded. However, as is well-known, condition (4.5) also holds for the standard normal density, the exponential density and more generally the density of the Gamma distribution, for example.

Since for every density $w(x)$ with support (a,b), $\int_a^b f(x)^2 dx < \infty$ implies that $f(x)/\sqrt{w(x)} \in L^2(w)$, the following corollary of Theorem 4.1 holds trivially.

Corollary 4.1. *Let $L^2(a,b)$, $-\infty \le a < b \le \infty$, be the Hilbert space of square integrable Borel measurable real functions on (a,b), with inner product $\langle f,g\rangle = \int_a^b f(x)g(x)dx$ and associated norm and metric. Every function $f \in L^2(a,b)$ can be written as*

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

where w is a density with support (a,b) satisfying the moment conditions (4.5), the $\overline{p}_k(x|w)$'s are the orthonormal polynomials generated by $w(x)$ and *the γ_k's are the Fourier coefficients of* $f(x)/\sqrt{w(x)}$, *i.e.,*

$$\gamma_k = \int_a^b f(x)\overline{p}_k(x|w)\sqrt{w(x)}dx.$$

This result implies that the functions

$$\psi_k\left(x|w\right) = \overline{p}_k(x|w)\sqrt{w(x)},\ k \in \mathbb{N},$$

form a complete orthonormal sequence in $L^2\left(a,b\right)$:

$$L^2\left(a,b\right) = \text{span}\left(\left\{\overline{p}_k(x|w)\sqrt{w(x)}\right\}_{k=0}^{\infty}\right).$$

Of course, the $\psi_k\left(x|w\right)$'s are no longer polynomials.

If $\max\left(|a|,|b|\right) < \infty$ then there is another way to construct a complete orthonormal sequence in $L^2\left(a,b\right)$, as follows. Let $W\left(x\right)$ be the distribution function of a density w with bounded support (a,b). Then

$$G\left(x\right) = a + \left(b-a\right)W\left(x\right)$$

is a one-to-one mapping of (a,b) onto (a,b), with inverse

$$G^{-1}(y) = W^{-1}\left(\left(y-a\right)/\left(b-a\right)\right)$$

where W^{-1} is the inverse of $W\left(x\right)$. For every $f \in L^2\left(a,b\right)$,

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

Hence $f\left(G\left(x\right)\right) \in L^2\left(w\right)$ and thus by Theorem 4.1,

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

where

$$\begin{aligned}\gamma_k &= \int_a^b f\left(G\left(x\right)\right)\overline{p}_k(x|w)w(x)dx = \frac{1}{b-a}\int_a^b f\left(G\left(x\right)\right)\overline{p}_k(x|w)dG(x) \\ &= \frac{1}{b-a}\int_a^b f\left(x\right)\overline{p}_k\left(G^{-1}\left(x\right)|w\right)dx\end{aligned}$$

Consequently

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

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

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

Thus,

Corollary 4.2. *Let w be a density with bounded support (a,b), satisfying the moment conditions (4.5). Let W be the c.d.f. of w, with inverse W^{-1}. Then the functions*

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

form a complete orthonormal sequence in $L^2(a,b)$, i.e., every $f \in L^2(a,b)$ can be written as $f(x) = \sum_{k=0}^{\infty} \alpha_k \psi_k(x|w)$ a.e. on (a,b), where $\alpha_k = \int_a^b f(x)$ $\psi_k(x|w)\,dx$.

4.2 The three-term recurrence relation

It follows from (4.1) that $p_0(x|w) \equiv 1$, and it follows from (4.2) that $p_1(x|w) = \alpha_{1,0} + x$ can be constructed by solving $\int_{-\infty}^{\infty} (\alpha_{1,0} + x)\,w(x)dx = 0$. Hence, given that $w(x)$ is a density, $\alpha_{1,0} = -\int_{-\infty}^{\infty} x.w(x)dx$. The question now arises how to construct these orthogonal polynomials further for $k \geq 2$.

The answer is the following.

Theorem 4.2. *Every sequence of polynomials* $p_k(x|w) = \sum_{j=0}^{k} \alpha_{k,j} x^j$*, with* $\alpha_{k,k} = 1$*, satisfying the orthogonality condition* (4.2)*, with* $w(x)$ *satisfying the moment conditions* (4.5)*, can be generated recursively by the three-term recurrence relation (hereafter referred to as TTRR)*

$$p_{k+1}(x|w) + (b_k - x)\,p_k(x|w) + c_k p_{k-1}(x|w) = 0,\ \ k \in \mathbb{N}, \tag{4.9}$$

where

$$b_k = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx} \tag{4.10}$$

and

$$c_k = \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx} \tag{4.11}$$

Next, let $d_k = \sqrt{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}$, so that $\overline{p}_k(x|w) = p_k(x|w)/d_k$ is a sequence of orthonormal polynomials. Substituting $p_k(x|w) = d_k.\overline{p}_k(x|w)$ in (4.9), (4.10) and (4.11) yields

$$\frac{d_{k+1}}{d_k}\overline{p}_{k+1}(x|w) + (b_k - x)\,\overline{p}_k(x|w) + c_k\frac{d_{k-1}}{d_k}\overline{p}_{k-1}(x|w) = 0,\ \ k \geq 1,$$

where $b_k = \int_{-\infty}^{\infty} x.\overline{p}_k(x|w)^2 w(x)dx$ and $c_k = d_k^2/d_{k-1}^2$, hence

$$\frac{d_{k+1}}{d_k}\overline{p}_{k+1}(x|w) + (b_k - x)\,\overline{p}_k(x|w) + \frac{d_k}{d_{k-1}}\overline{p}_{k-1}(x|w) = 0,\ \ k \geq 1.$$

Moreover, note that

$$\lim_{|x|\to\infty} \frac{x\overline{p}_{k-1}(x|w)}{\overline{p}_k(x|w)} = \frac{d_k}{d_{k-1}} \lim_{|x|\to\infty} \frac{x.p_{k-1}(x|w)}{p_k(x|w)} = \frac{d_k}{d_{k-1}},$$

where the latter equality is due to the normalization $\alpha_{k,k} = 1$ in Theorem 4.2. Thus:

Theorem 4.3. *Every sequence* $\overline{p}_k(x|w)$ *of orthonormal polynomials with respect to a density function* $w(x)$ *satisfying the moment conditions* (4.5) *can be generated recursively by the TTRR*

$$a_{k+1}.\overline{p}_{k+1}(x|w) + (b_k - x)\,\overline{p}_k(x|w) + a_k.\overline{p}_{k-1}(x|w) = 0,\ \ \ k \in \mathbb{N}. \tag{4.12}$$

where

$$a_k = \left| \lim_{|x|\to\infty} \frac{x.\overline{p}_{k-1}(x|w)}{\overline{p}_k(x|w)} \right| \tag{4.13}$$

and

$$b_k = \int_{-\infty}^{\infty} x.\overline{p}_k(x|w)^2 w(x)dx. \tag{4.14}$$

4.3 Examples of orthonormal polynomials

4.3.1 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

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

starting from $\overline{p}_0(x|w_{\mathcal{N}[0,1]}) = 1$, $\overline{p}_1(x|w_{\mathcal{N}[0,1]}) = x$. Thus in this case $a_k = \sqrt{k}$ and $b_k = 0$ in (4.12). These polynomials are known as Hermite[1] polynomials.

It follows from Theorem 4.1 that the Hermite polynomials span the Hilbert space $L^2(w_{\mathcal{N}[0,1]})$, and it follows from Corollary 4.1 that

$$L^2(\mathbb{R}) = \text{ span}\left(\left\{\sqrt{w_{\mathcal{N}[0,1]}(x)}\overline{p}_k(x|w_{\mathcal{N}[0,1]})\right\}_{k=0}^{\infty}\right).$$

Consequently, any density $f(x)$ on $\mathbb{R}$ can be represented by

$$f(x) = w_{\mathcal{N}[0,1]}\left(x\right)\left(\sum_{k=0}^{\infty}\gamma_k\overline{p}_k(x|w_{\mathcal{N}[0,1]})\right)^2$$

where $\sum_{k=0}^{\infty}\gamma_k^2 = 1$.

[1] Charles Hermite (1822-1901).

4.3.2 Laguerre polynomials

The standard exponential density function

$$w_{\mathrm{Exp}}(x) = I\left(x \geq 0\right)\exp\left(-x\right)$$

gives rise to the orthonormal Laguerre[2] polynomials, with TTRR

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

for $k \in \mathbb{N}$, starting from $\overline{p}_0(x|w_{\mathrm{Exp}}) = 1$, $\overline{p}_1(x|w_{\mathrm{Exp}}) = x - 1$. Thus in this case $a_k = k$ and $b_k = 2k + 1$.

Since the moment conditions (4.5) hold for $w_{\mathrm{Exp}}(x)$, it follows from Theorem 4.1 that any Borel measurable real function $f(x)$ satisfying $\int_0^\infty \exp(-x)$ $f(x)^2 dx < \infty$ can be written as $f(x) = \sum_{k=0}^\infty \gamma_k \overline{p}_k(x|w_{\mathrm{Exp}})$ a.e. on $[0,\infty)$, where $\gamma_k = \int_0^\infty \exp(-x)\overline{p}_{k+1}(x|w)f(x)dx$.

Again, it follows from Corollary 4.1 that

$$L^2(0,\infty) = \mathrm{span}\left(\{\exp(-x/2)\overline{p}_k(x|w_{\mathrm{Exp}})\}_{k=0}^\infty\right),$$

hence any density $f(x)$ on $[0,\infty)$ can be written as

$$f(x) = \exp(-x)\left(\sum_{k=0}^\infty \gamma_k \overline{p}_k(x|w_{\mathrm{Exp}})\right)^2, \text{ with } \sum_{k=0}^\infty \gamma_k^2 = 1.$$

4.3.3 Legendre polynomials

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

$$w_{\mathcal{U}[-1,1]}(x) = \frac{1}{2}I\left(|x| \leq 1\right),$$

generates the orthonormal Legendre[3] polynomials on $[-1,1]$, with TTRR

$$\frac{k+1}{\sqrt{2k+3}\sqrt{2k+1}}\overline{p}_{k+1}(x|w_{\mathcal{U}[-1,1]}) - x.\overline{p}_k(x|w_{\mathcal{U}[-1,1]}) + \frac{k}{\sqrt{2k+1}\sqrt{2k-1}}\overline{p}_{k-1}(x|w_{\mathcal{U}[-1,1]}) = 0, \tag{4.15}$$

[2] Edmund Nicolas Laguerre (1834-1886)
[3] Adrien-Marie Legendre (1752-1833)

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

Note that the orthonormal Legendre polynomials $\overline{p}_k(x|w_{\mathcal{U}[-1,1]})$ satisfy

$$\begin{aligned}
&\int_0^1 \overline{p}_k(2u-1|w_{\mathcal{U}[-1,1]})\overline{p}_m(2u-1|w_{\mathcal{U}[-1,1]})du \\
&\qquad = \frac{1}{2}\int_0^1 \overline{p}_k(2u-1|w_{\mathcal{U}[-1,1]})\overline{p}_m(2u-1|w_{\mathcal{U}[-1,1]})d\left(2u-1\right) \\
&\qquad = \frac{1}{2}\int_{-1}^1 \overline{p}_k(x|w_{\mathcal{U}[-1,1]})\overline{p}_m(x|w_{\mathcal{U}[-1,1]})dx = I\left(k=m\right)
\end{aligned}$$

Hence,

$$\overline{p}_k(u|w_{\mathcal{U}[0,1]}) = \overline{p}_k(2u-1|w_{\mathcal{U}[-1,1]}),\ \ k \in \mathbb{N}_0,$$

is a sequence of orthonormal polynomials w.r.t. the uniform density on $[0,1]$,

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

The $\overline{p}_k(u|w_{\mathcal{U}[0,1]})$'s are known as the shifted Legendre polynomials, also called the orthonormal Legendre polynomials on the unit interval $[0,1]$. Substituting $x = 2u-1$ and $\overline{p}_k(x|w_{\mathcal{U}[-1,1]})) = \overline{p}_k(u|w_{\mathcal{U}[0,1]})$ in (4.15) yields the TTRR

$$\begin{aligned}
&\frac{\left(k+1\right)/2}{\sqrt{2k+3}\sqrt{2k+1}}\rho_{k+1}(u|w_{\mathcal{U}[0,1]}) + \left(0.5-u\right).\rho_k(u|w_{\mathcal{U}[0,1]}) \\
&\qquad + \frac{k/2}{\sqrt{2k+1}\sqrt{2k-1}}\rho_{k-1}(u|w_{\mathcal{U}[0,1]}) = 0,\ \ k \in \mathbb{N},
\end{aligned}$$

starting from $\rho_0(u) = 1$, $\rho_1(u) = \sqrt{3}\left(2u-1\right)$.

Again, it follows from Theorem 4.1 that any Borel measurable real function $f(x)$ on $[0,1]$ can be written as $f(x) = \sum_{k=0}^{\infty}\gamma_k\rho_k(u|w_{\mathcal{U}[0,1]})$, where $\gamma_k = \int_0^1 f\left(x\right)\rho_k(u|w_{\mathcal{U}[0,1]})dx$, hence $L^2(0,1) = \text{span}\left(\left\{\rho_k(u|w_{\mathcal{U}[0,1]}\right\}_{k=0}^{\infty}\right)$.

These shifted Legendre polynomials have been used by Bierens (2008) and Bierens and Carvalho (2007) to model semi-nonparametrically the unobserved heterogeneity of interval-censored mixed proportional hazard models and bivariate mixed proportional hazard models, respectively.

4.3.4 Chebyshev polynomials

Chebyshev polynomials on $[-1, 1]$

Chebyshev polynomials on $[-1, 1]$ are generated by the weight function

$$w_{\mathcal{C}[-1,1]}(x) = \frac{1}{\pi\sqrt{1-x^2}} I\left(|x| < 1\right). \tag{4.16}$$

This is a density function on $(-1, 1)$. To see this, let $\theta = \arccos(x)$, so that $x = \cos(\theta)$, and observe that

$$\frac{dx}{d\theta} = -\sin(\theta) = -\sqrt{1-\cos^2(\theta)} = -\sqrt{1-x^2},$$

hence

$$\frac{d\arccos(x)}{dx} = \frac{-1}{\sqrt{1-x^2}} \tag{4.17}$$

Then

$$\begin{aligned} \int_{-1}^{1} \frac{1}{\pi\sqrt{1-x^2}} dx &= -\frac{1}{\pi} \int_{-1}^{1} d\arccos(x) \\ &= \frac{\arccos(-1) - \arccos(1)}{\pi} = 1 \end{aligned}$$

because $\arccos(-1) = \pi$ and $\arccos(1) = 0$. Clearly, the corresponding distribution function is

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

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

$$p_{k+1}(x|w_{\mathcal{C}[-1,1]}) - 2xp_k(x|w_{\mathcal{C}[-1,1]}) + p_{k-1}(x|w_{\mathcal{C}[-1,1]}) = 0, \ k \in \mathbb{N}, \tag{4.18}$$

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

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

An important practical difference with the other polynomials discussed so far is that Chebyshev polynomials have the closed form[4]:

$$p_k(x|w_{\mathcal{C}[-1,1]}) = \cos\left(k.\arccos(x)\right). \tag{4.19}$$

To see this, observe from (4.17) and the well-known sine-cosine formulas that

$$\begin{aligned} &\int_{-1}^{1} \frac{\cos\left(k.\arccos(x)\right)\cos\left(m.\arccos(x)\right)}{\pi\sqrt{1-x^2}}dx \\ &= -\frac{1}{\pi}\int_{-1}^{1} \cos\left(k.\arccos(x)\right)\cos\left(m.\arccos(x)\right) d\arccos(x) \\ &= \frac{1}{\pi}\int_{0}^{\pi} \cos\left(k.\theta\right)\cos\left(m.\theta\right) d\theta \\ &= \frac{1}{2\pi}\int_{0}^{\pi} \cos\left(\left(k+m\right)\theta\right) d\theta + \frac{1}{2\pi}\int_{0}^{\pi} \cos\left(\left(k-m\right)\theta\right) d\theta \\ &= \frac{1}{2}\left(\frac{\sin\left(\left(k+m\right)\pi\right)}{\left(k+m\right)\pi} + \frac{\sin\left(\left(k-m\right)\pi\right)}{\left(k-m\right)\pi}\right) \\ &= \begin{cases} 0 & \text{if } \ k \neq m, \\ 1/2 & \text{if } \ k = m > 0, \\ 1 & \text{if } \ k = m = 0. \end{cases} \end{aligned} \tag{4.20}$$

Moreover, the TTRR (4.18) follows from

$$\begin{aligned} &\cos\left(\left(k+1\right).\theta\right) - 2\cos(\theta)\cos\left(k.\theta\right) + \cos\left(\left(k-1\right).\theta\right) \\ &= \cos\left(k.\theta\right)\cos(\theta) - \sin\left(k.\theta\right)\sin(\theta) - 2\cos(\theta)\cos\left(k.\theta\right) \\ &+ \cos\left(k.\theta\right)\cos(\theta) + \sin\left(k.\theta\right)\sin(\theta) = 0. \end{aligned}$$

Hence, the functions (4.19) satisfy the TTRR (4.18) and are therefore genuine polynomials.

In view of (4.20) we can now define the orthonormal Chebyshev polynomials as

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

[4]Note that $\arccos(x) = \operatorname{atan}\left(-x/\sqrt{1-x^2}\right) + \frac{1}{2}.\pi$, where $\operatorname{atan}(x)$ is the inverse of the tangents function $\tan(\theta) = \sin(\theta)/\cos(\theta)$, $\theta \in (-\pi/2, \pi/2)$. In most programming languages the function $\operatorname{atan}(x)$ is an intrinsic function. For example, in Visual Basic this function is the ATN(x) function.

It is trivial to verify that the density (4.16) satisfies the moment condition (4.5), so that the Chebyshev polynomials form a complete orthonormal sequence in the Hilbert space $L^2(w_{\mathcal{C}[-1,1]})$ involved.

Further properties of Chebyshev polynomials

Because $p_n(x|w_{\mathcal{C}[-1,1]})$ is a polynomial of order n in $x \in [-1,1]$, it has at most n real roots in $[-1,1]$. Obviously, these roots are

$$x_{n,k} = \cos\left(\pi(k-0.5)/n\right),\ \ k = 1,2,...,n$$

Moreover,

Lemma 4.1. *For* $j_1, j_2 = 0,1,2,...,n-1$,

$$\begin{aligned} &\sum_{k=1}^{n} \cos\left(\pi j_1(k-0.5)/n\right)\cos\left(\pi j_2(k-0.5)/n\right) \\ &= \sum_{k=1}^{n} p_{j_1}(x_{n,k}|w_{\mathcal{C}[-1,1]})p_{j_2}(x_{n,k}|w_{\mathcal{C}[-1,1]}) = \begin{cases} 0 & if \quad j_1 \neq j_2, \\ n/2 & if \quad j_1 = j_2 > 0, \\ n & if \quad j_1 = j_2 = 0. \end{cases} \end{aligned}$$

Now interpret k in Lemma 4.1 as a time index: $k = t = 1, ..., n$, and denote

$$\begin{aligned} P_{0,n}(t) &\equiv 1,\ P_{j,n}(t) = \sqrt{2}\cos(j\pi(t-0.5)/n), \\ j &= 1,2,...,n-1,\ t = 1,2,....,n. \end{aligned}$$

The $P_{j,n}(t)$'s are known as Chebyshev time polynomials, which by Lemma 4.1 satisfy

$$\frac{1}{n}\sum_{t=1}^{n} P_{i,n}(t)P_{j,n}(t) = I(i=j),\ i,j = 0,1,2,...,n-1.$$

Consequently, any function $g(t)$ of time $t = 1,2,...,n$ can be represented by

$$g(t) = \sum_{j=1}^{n-1} c_{j,n}P_{j,n}(t),\ \text{where } c_{j,n} = \frac{1}{n}\sum_{k=1}^{n} g(k)P_{j,n}(k).$$

In particular, if $g(t)$ is smooth then

$$g(t) \approx \sum_{j=1}^{m} c_{j,n} P_{j,n}(t)$$

for modest values of m. This approximation has been used in Bierens (1997) to test the unit root hypothesis against nonlinear trend stationarity, and in Bierens and Martins (2010) to test for time varying cointegration.

Shifted Chebyshev polynomials

Substituting $x = 2u - 1$ for $u \in [0,1]$ in (4.16) yields

$$w_{\mathcal{C}[0,1]}(u) = \frac{2}{\pi\sqrt{1-(2u-1)^2}} = \frac{1}{\pi\sqrt{u(1-u)}}. \tag{4.21}$$

with corresponding distribution function

$$W_{\mathcal{C}[0,1]}(u) = 1 - \pi^{-1}\arccos(2u-1), \tag{4.22}$$

and shifted Chebyshev polynomials

$$\overline{p}_k(u|w_{\mathcal{C}[0,1]}) = \begin{cases} 1 & \text{for} \quad k=0, \\ \sqrt{2}\cos\left(k.\arccos(2u-1)\right) & \text{for} \quad k \in \mathbb{N}. \end{cases} \tag{4.23}$$

Again, it follows from Corollary 4.1 that the orthonormal sequence

$$\psi_k(u) = \begin{cases} \sqrt{w_{\mathcal{C}[0,1]}(u)} & \text{for} \quad k=0, \\ \sqrt{2}\sqrt{w_{\mathcal{C}[0,1]}(u)}\cos\left(k.\arccos(2u-1)\right) & \text{for} \quad k \in \mathbb{N}, \end{cases}$$

is complete in $L^2(0,1)$. Thus, every function $f \in L^2(0,1)$ can be written as

$$f(u) = \sum_{k=0}^{\infty} \gamma_k \psi_k(u) \tag{4.24}$$

a.e. on $(0,1)$, where $\gamma_k = \int_0^1 f(u)\,\psi_k(u)du$.

4.4 Bivariate functions

Let $w_1(x)$ and $w_2(y)$ be densities with supports (a_1, b_1) and (a_2, b_2), respectively, where $\infty \leq a_i < b_i \leq \infty$, $i = 1, 2$, satisfying the conditions of Theorem 4.1. Consider the space $L^2(w_1 \times w_2)$ of bivariate Borel measurable functions $f(x, y)$ satisfying

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} w_1(x) w_2(y) f(x, y)^2 dx dy < \infty, \tag{4.25}$$

endowed with the inner product

$$\langle f, g \rangle = \int_{a_1}^{b_1} \int_{a_2}^{b_2} w_1(x) w_2(y) f(x, y) g(x, y) dx dy$$

and associated norm $||f|| = \sqrt{\langle f, f \rangle}$ and metric $||f - g||$. Then for any fixed $y \in (a_2, b_2)$ for which

$$\int_{a_1}^{b_1} w_1(x) f(x, y)^2 dx < \infty, \tag{4.26}$$

we have $f(x, y) \in L^2(w_1)$, hence

$$f(x, y) = \sum_{k=0}^{\infty} \gamma_k(y) \overline{p}_k(x|w_1) \text{ a.e. on } (a_1, b_1). \tag{4.27}$$

where $\gamma_k(y) = \int_{a_1}^{b_1} w_1(x) f(x, y) \overline{p}_k(x|w_1) dx$ and $\sum_{k=0}^{\infty} \gamma_k(y)^2 < \infty$.

Note that by the Cauchy-Schwarz inequality and (4.25),

$$\begin{aligned}
\int_{a_2}^{b_2} w_2(y) \gamma_k(y)^2 dy &= \int_{a_2}^{b_2} w_2(y) \left(\int_{a_1}^{b_1} w_1(x) \overline{p}_k(x|w_1) f(x, y) dx \right)^2 dy \\
&\leq \int_{a_1}^{b_1} w_1(x) \overline{p}_k(x|w_1)^2 dx \int_{a_1}^{b_1} \int_{a_2}^{b_2} w_1(x) w_2(y) f(x, y)^2 dx dy \\
&= \int_{a_1}^{b_1} \int_{a_2}^{b_2} w_1(x) w_2(y) f(x, y)^2 dx dy < \infty
\end{aligned}$$

where the second equality follows from the fact that $\int_{a_1}^{b_1} w_1(x) \overline{p}_k(x|w_1)^2 dx = 1$, so that $\gamma_k(y) \in L^2(w_2)$. Consequently, for each $k \in \mathbb{N}_0$ we have

$$\gamma_k(y) = \sum_{m=0}^{\infty} \gamma_{k,m} \overline{p}_m(y|w_2) \text{ a.e. on } (a_2, b_2), \tag{4.28}$$

where

$$\begin{aligned}\gamma_{k,m} &= \int_{a_2}^{b_2} w_2(y)\gamma_k(y)\overline{p}_m(y|w_2)dy \\ &= \int_{a_1}^{b_1}\int_{a_2}^{b_2} w_1(x)w_2(y)f(x,y)\overline{p}_k(x|w_1)\overline{p}_m(y|w_2)dxdy \qquad (4.29)\end{aligned}$$

and $\sum_{m=0}^{\infty}\gamma_{k,m}^2 < \infty$.

Moreover, note that due to (4.25) the restriction (4.26) holds a.e. on (a_2, b_2), so that

$$f(x,y) = \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\gamma_{k,m}\overline{p}_k(x|w_1)\overline{p}_m(y|w_2) \text{ a.e. on } (a_1,b_1)\times(a_2,b_2),$$

where the double-array $\gamma_{k,m}$ of Fourier coefficients are given by (4.29) and satisfies $\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\gamma_{k,m}^2 < \infty$. Consequently, the space $L^2(w_1 \times w_2)$ is a Hilbert space.

Recall that in the case

$$w_1(x) = w_2(x) = \exp(-x^2/2)/\sqrt{2\pi} = w_{\mathcal{N}[0,1]}(x)$$

the polynomials $\overline{p}_k(x|w_{\mathcal{N}[0,1]})$ are the Hermite polynomials. Then every density $f(x,y)$ on $\mathbb{R}^2$ can be written as

$$\begin{aligned}f(x,y) &= \frac{\exp\left(-\frac{1}{2}\left(x^2+y^2\right)\right)}{2\pi} \\ &\quad \times \left(\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\gamma_{k,m}\overline{p}_k(x|w_{\mathcal{N}[0,1]})\overline{p}_m(y|w_{\mathcal{N}[0,1]})\right)^2 \qquad (4.30) \\ &\quad \text{a.e. on } \mathbb{R}^2, \text{ where } \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\gamma_{k,m}^2 = 1.\end{aligned}$$

This is the approach taken by Gallant and Nychka (1987). They consider SNP estimation of Heckman's (1979) sample selection model, where the bivariate error distribution of the latent variable equations involved is modeled semi-nonparametrically via the Hermite expansion (4.30) of the error density.

4.5 Appendix: Proofs

4.5.1 Theorem 4.1

Let $\overline{f}_n(x) = \sum_{k=0}^n \gamma_k \overline{p}_k(x|w)$, where $\gamma_k = \int_a^b \overline{p}_k(x|w)f(x)w(x)dx$, and observe that due to condition (4.5), $\overline{f}_n \in L^2(w)$. Next, observe that

$$\begin{aligned}
\left\|f - \overline{f}_n\right\|^2 &= \int_a^b \left(f(x) - \sum_{k=0}^n \gamma_k \overline{p}_k(x|w)\right)^2 w(x)dx \\
&= \int_a^b f(x)^2 w(x)dx - 2\sum_{k=0}^n \gamma_k \int_a^b \overline{p}_k(x|w)f(x)w(x)dx \\
&\quad + \sum_{k_1=0}^n \sum_{k_1=0}^n \gamma_{k_1}\gamma_{k_2} \int_a^b \overline{p}_{k_1}(x|w)\overline{p}_{k_2}(x|w)w(x)dx \\
&= \int_a^b f(x)^2 w(x)dx - \sum_{k=0}^n \gamma_k^2 \geq 0. \qquad (4.31)
\end{aligned}$$

Hence $\sum_{k=0}^n \gamma_k^2 \leq \int_a^b f(x)^2 w(x)dx < \infty$ for all $n \geq 0$, and thus

$$\sum_{k=0}^\infty \gamma_k^2 < \infty. \qquad (4.32)$$

The latter implies that $\overline{f}_n$ is a Cauchy sequence in $L^2(w)$ because

$$\lim_{\min(n,m)\to\infty} \left\|\overline{f}_n - \overline{f}_m\right\|^2 = \lim_{\min(n,m)\to\infty} \sum_{k=\min(n,m)+1}^{\max(n,m)} \gamma_k^2 \leq \lim_{m\to\infty} \sum_{k=m}^\infty \gamma_k^2 = 0.$$

Therefore, there exists a function $\overline{f} \in L^2(w)$ such that

$$\lim_{n\to\infty} \left\|\overline{f}_n - \overline{f}\right\| = 0. \qquad (4.33)$$

This limit function $\overline{f}$ can be written as

$$\overline{f}(x) = \sum_{k=0}^n \gamma_k \overline{p}_k(x|w) + \varepsilon_n(x) \qquad (4.34)$$

for all $n \in \mathbb{N}$, where

$$\lim_{n\to\infty} \int_a^b \varepsilon_n(x)^2 w(x)dx = 0. \tag{4.35}$$

Proof of (4.7)
To prove (4.7), it suffices to show that

$$\int_a^b \exp(\mathbf{i}.t.x)\left(f(x) - \overline{f}(x)\right) w(x)dx = 0 \tag{4.36}$$

for all $t \in R$, because (4.36) implies that $f(x) = \overline{f}(x)$ a.e. on (a,b), due to the uniqueness of the Fourier transform.[5] .

It follows from the definition of γ_m and $\overline{f}$ that for $m \leq n$,

$$\begin{aligned}\left|\int_a^b \left(f(x) - \overline{f}(x)\right)\overline{p}_m(x|w)w(x)dx\right| &= \left|\int_a^b \varepsilon_n(x)\overline{p}_m(x|w)w(x)dx\right| \\ &\leq \sqrt{\int_a^b \varepsilon_n(x)^2 w(x)dx},\end{aligned}$$

hence by (4.35),

$$\int_a^b \left(f(x) - \overline{f}(x)\right)\overline{p}_m(x|w)w(x)dx = 0 \tag{4.37}$$

for all $m \in \mathbb{N}$. This result implies, by induction, that

$$\int_a^b \left(f(x) - \overline{f}(x)\right)x^m w(x)dx = 0 \text{ for all } m \in \mathbb{N}. \tag{4.38}$$

In its turn (4.38) implies, together with the well-known equality $\exp(\mathbf{i}.t.x) = \sum_{m=0}^{\infty}(\mathbf{i}.t.x)^m/m!$, that for $t \in \mathbb{R}$ and all $n \in \mathbb{N}$,

$$\begin{aligned}&\int_a^b \exp(\mathbf{i}.t.x)\left(f(x) - \overline{f}(x)\right)w(x)dx \\ &\quad= \int_a^b \sum_{m=0}^{n} \frac{(\mathbf{i}.t.x)^m}{m!}\left(f(x) - \overline{f}(x)\right)w(x)dx\end{aligned}$$

[5] See for example Bierens (1994, Theorem 3.1.1, p.50).

$$+\int_a^b \left(\sum_{m=n+1}^{\infty} \frac{(\mathbf{i}.t.x)^m}{m!} \right) \left(f(x) - \overline{f}(x) \right) w(x)dx$$
$$= \int_a^b \left(\sum_{m=n+1}^{\infty} \frac{(\mathbf{i}.t.x)^m}{m!} \right) \left(f(x) - \overline{f}(x) \right) w(x)dx$$

If $-\infty < a < b < \infty$ then by dominated convergence,

$$\int_a^b \exp(\mathbf{i}.t.x) \left(f(x) - \overline{f}(x) \right) w(x)dx$$
$$= \int_a^b \left(\lim_{n\to\infty} \sum_{m=n+1}^{\infty} \frac{(\mathbf{i}.t.x)^m}{m!} \right) \left(f(x) - \overline{f}(x) \right) w(x)dx = 0$$

If $a = -\infty$ and/or $b = \infty$ we can find for arbitrary $\varepsilon > 0$ a finite lower bound $a(\varepsilon) > a$ and a finite upper bound $b(\varepsilon) < b$ such that

$$\left| \int_a^{a(\varepsilon)} \exp(\mathbf{i}.t.x) \left(f(x) - \overline{f}(x) \right) w(x)dx \right| < \varepsilon/2$$
$$\left| \int_{b(\varepsilon)}^b \exp(\mathbf{i}.t.x) \left(f(x) - \overline{f}(x) \right) w(x)dx \right| < \varepsilon/2$$

whereas by dominated convergence

$$\int_{a(\varepsilon)}^{b(\varepsilon)} \exp(\mathbf{i}.t.x) \left(f(x) - \overline{f}(x) \right) w(x)dx$$
$$= \int_{a(\varepsilon)}^{b(\varepsilon)} \left(\lim_{n\to\infty} \sum_{m=n+1}^{\infty} \frac{(\mathbf{i}.t.x)^m}{m!} \right) \left(f(x) - \overline{f}(x) \right) w(x)dx = 0$$

Since $\varepsilon > 0$ is arbitrary, we therefore have in either case that (4.36) holds. It therefore follows from (4.34) and (4.35) that

$$\lim_{n\to\infty} \int_a^b \left(f(x) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(x|w) \right)^2 w(x)\,dx = 0. \tag{4.39}$$

This completes the proof of (4.7).

Proof of (4.8)

To prove that (4.7) implies (4.8), let X be a random drawing from $w(x)$. Then by Chebyshev's inequality, (4.39) implies

$$f\left(X\right)=\underset{n\rightarrow\infty}{\text{plim}}\sum_{k=0}^{n}\gamma_{k}\overline{p}_{k}(X|w) \tag{4.40}$$

As is well-known[6], convergence in probability is equivalent to almost sure (a.s.) convergence along a further subsequence of an arbitrary subsequence of n. Thus it follows from (4.40) that for any subsequence n_j in $\mathbb{N}$ there exists a further subsequence n_{j_m} such that for $m\rightarrow\infty$,

$$\sum_{k=0}^{n_{j_m}}\gamma_{k}\overline{p}_{k}(X|w)\overset{a.s.}{\rightarrow}f(X). \tag{4.41}$$

For each n there exists an m such that $n_{j_{m-1}}\leq n<n_{j_m}$. Hence, there exists a further subsequence j_n of n_{j_m} such that for $j_{n-1}\leq n<j_n$ and $n\rightarrow\infty$,

$$\sum_{k=0}^{j_n}\gamma_{k}\overline{p}_{k}(X|w)\overset{a.s.}{\rightarrow}f(X). \tag{4.42}$$

The latter implies that

$$\begin{aligned}E\left[\left(\sum_{k=0}^{j_n}\gamma_{k}\overline{p}_{k}(X|w)-\sum_{k=0}^{n}\gamma_{k}\overline{p}_{k}(X|w)\right)^{2}\right] &= E\left(\sum_{k=n+1}^{j_n}\gamma_{k}\overline{p}_{k}(X|w)\right)^{2}\\ &\leq \sum_{k=j_{n-1}+1}^{j_n}\gamma_{k}^{2}\end{aligned}$$

so that

$$\begin{aligned}\sum_{n=1}^{\infty}E\left[\left(\sum_{k=0}^{j_n}\gamma_{k}\overline{p}_{k}(X|w)-\sum_{k=0}^{n}\gamma_{k}\overline{p}_{k}(X|w)\right)^{2}\right] &\leq \sum_{n=1}^{\infty}\sum_{k=j_{n-1}+1}^{j_n}\gamma_{k}^{2}\\ &\leq \sum_{k=0}^{\infty}\gamma_{k}^{2}<\infty\end{aligned}$$

[6] See for example Bierens (2004, Theorem 6.B.3, p.168).

Then by Chebyshev's inequality,

$$\sum_{n=1}^{\infty} \Pr\left[\left|\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w)\right| > \varepsilon\right] < \infty$$

for all $\varepsilon > 0$, which by the Borel-Cantelli lemma[7] implies that for $n \to \infty$

$$\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w) \overset{a.s.}{\to} 0. \tag{4.43}$$

Combining (4.42) and (4.43), it follows that $\sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w) \overset{a.s.}{\to} f(X)$ as $n \to \infty$, which is equivalent to (4.8) because the support of $w(x)$ was assumed to be (a, b).

4.5.2 Theorem 4.2

Due to the normalization $\alpha_{k,k} = 1$ it follows that $p_{k+1}(x|w) - x.p_k(x|w)$ is a polynomial of order k, which can be written as a linear combination of $p_0(x|w), p_1(x|w),...,p_k(x|w)$:

$$p_{k+1}(x|w) - x.p_k(x|w) = \sum_{j=0}^{k} \delta_{j,k} p_j(x|w) \tag{4.44}$$

for example. Then for $m \leq k$,

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} p_{k+1}(x|w)p_m(x|w)w(x)dx - \int_{-\infty}^{\infty} x.p_k(x|w)p_m(x|w)w(x)dx \\ &\quad - \sum_{j=0}^{k} \delta_{j,k} \int_{-\infty}^{\infty} p_j(x|w)p_m(x|w)w(x)dx \\ &= -\int_{-\infty}^{\infty} x.p_k(x|w)p_m(x|w)w(x)dx - \delta_{m,k} \int_{-\infty}^{\infty} p_m(x|w)^2 w(x)dx \end{aligned}$$

so that

$$\delta_{m,k} = -\frac{\int_{-\infty}^{\infty} (x.p_m(x|w))\, p_k(x|w)w(x)dx}{\int_{-\infty}^{\infty} p_m(x|w)^2 w(x)dx}, \; m = 0, 1, 2, ..., k.$$

[7] See for example Bierens (2004, Theorem 2.B.2, p. 168).

Because $x.p_m(x|w)$ is a polynomial of order $m+1$, it follows that for $m \leq k-2$, $x.p_m(x|w)$ is orthogonal to $p_k(x|w)$, hence $\delta_{m,k}=0$ for $m = 0,1,....,k-2$. Thus it follows from (4.44) that

$$\begin{aligned} p_{k+1}(x|w) - x.p_k(x|w) &= \delta_{k,k}p_k(x|w) + \delta_{k-1,k}p_{k-1}(x|w) \\ &= -b_k p_k(x|w) - c_k p_{k-1}(x|w) \end{aligned}$$

where

$$b_k = -\delta_{k,k} = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}$$

and

$$\begin{aligned} c_k = -\delta_{k-1,k} &= \frac{\int_{-\infty}^{\infty} x.p_{k-1}(x|w).p_k(x|w)w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx} \\ &= \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx} \end{aligned}$$

The last equality follows from the fact that $x.p_{k-1}(x|w)$ can be written as $x.p_{k-1}(x|w) = \sum_{m=0}^{k} \beta_{m,k} p_m(x|w)$, where $\beta_{k,k}=1$, so that

$$\begin{aligned} \int_{-\infty}^{\infty} x.p_{k-1}(x|w).p_k(x|w)w(x)dx &= \sum_{m=0}^{k} \beta_{m,k} \int_{-\infty}^{\infty} p_m(x|w)p_k(x|w)w(x)dx \\ &= \beta_{k,k} \int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx \\ &= \int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx. \end{aligned}$$

4.5.3 Lemma 4.1

Using the well-known cosine formulas $2\cos(a)\cos(b) = \cos(a+b)+\cos(a-b)$ and $\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b)$ we can write

$$\begin{aligned} &\sum_{k=1}^{n} p_{j_1}(x_{n,k}|w_{\mathcal{C}[-1,1]}) p_{j_2}(x_{n,k}|w_{\mathcal{C}[-1,1]}) \\ &= \sum_{k=1}^{n} \cos\left(\pi j_1(k-0.5)/n\right) \cos\left(\pi j_2(k-0.5)/n\right) \end{aligned}$$

$$= \frac{1}{2}\sum_{k=1}^{n}\cos\left(\pi(j_1+j_2)(k-0.5)/n\right) + \frac{1}{2}\sum_{k=1}^{n}\cos\left(\pi(j_1-j_2)(k-0.5)/n\right)$$
$$= \frac{1}{2}\cos\left(0.5\pi(j_1+j_2)/n\right)\sum_{k=1}^{n}\cos\left(\pi(j_1+j_2)k/n\right)$$
$$+\frac{1}{2}\sin\left(0.5\pi(j_1+j_2)/n\right)\sum_{k=1}^{n}\sin\left(\pi(j_1+j_2)k/n\right)$$
$$+\frac{1}{2}\cos\left(0.5\pi(j_1-j_2)/n\right)\sum_{k=1}^{n}\cos\left(\pi(j_1-j_2)k/n\right)$$
$$+\frac{1}{2}\sin\left(0.5\pi(j_1-j_2)/n\right)\sum_{k=1}^{n}\sin\left(\pi(j_1-j_2)k/n\right)$$

Moreover, using the well-known De Moivre formula $\exp(\mathbf{i}.a) = \cos(a) + \mathbf{i}.\sin(a)$ it follows that

$$\frac{1}{2}\sum_{k=1}^{n}\cos\left(\pi.m.k/n\right)$$
$$= \sum_{k=1}^{n}\exp\left(\mathbf{i}.\pi m.k/n\right) + \sum_{k=1}^{n}\exp\left(-\mathbf{i}.\pi m.k/n\right)$$
$$= \sum_{k=1}^{n}\left(\exp\left(\mathbf{i}.\pi m/n\right)\right)^k + \sum_{k=1}^{n}\left(\exp\left(-\mathbf{i}.\pi m/n\right)\right)^k$$
$$= \exp\left(\mathbf{i}.\pi m/n\right)\frac{\exp\left(\mathbf{i}.\pi m\right)-1}{\exp\left(\mathbf{i}.\pi m/n\right)-1} + \exp\left(-\mathbf{i}.\pi m/n\right)\frac{\exp\left(-\mathbf{i}.\pi m\right)-1}{\exp\left(-\mathbf{i}.\pi m/n\right)-1}$$
$$= \frac{\exp\left(\mathbf{i}.\pi m/(2n)\right)}{\exp\left(\mathbf{i}.\pi m/(2n)\right)-\exp\left(-\mathbf{i}.\pi m/(2n)\right)}\left(\cos\left(\pi m\right)-1\right)$$
$$-\frac{\exp\left(-\mathbf{i}.\pi m/(2n)\right)}{\exp\left(\mathbf{i}.\pi m/(2n)\right)-\exp\left(-\mathbf{i}.\pi m/(2n)\right)}\left(\cos\left(\pi m\right)-1\right)$$
$$= \cos\left(\pi m\right)-1$$

and similarly for $m \neq 0$,

$$\frac{1}{2}\sum_{k=1}^{n}\sin\left(\pi.m.k/n\right)$$

$$
\begin{aligned}
&= \frac{1}{\mathbf{i}}\sum_{k=1}^{n}\exp\left(\mathbf{i}.\pi m.k/n\right) - \frac{1}{\mathbf{i}}\sum_{k=1}^{n}\exp\left(-\mathbf{i}.\pi m.k/n\right)\\
&= \frac{1}{\mathbf{i}}\sum_{k=1}^{n}\left(\exp\left(\mathbf{i}.\pi m/n\right)\right)^k - \frac{1}{\mathbf{i}}\sum_{k=1}^{n}\left(\exp\left(-\mathbf{i}.\pi m/n\right)\right)^k\\
&= \frac{1}{\mathbf{i}}\exp\left(\mathbf{i}.\pi m/n\right)\frac{\exp\left(\mathbf{i}.\pi m\right)-1}{\exp\left(\mathbf{i}.\pi m/n\right)-1} - \frac{1}{\mathbf{i}}\exp\left(-\mathbf{i}.\pi m/n\right)\frac{\exp\left(-\mathbf{i}.\pi m\right)-1}{\exp\left(-\mathbf{i}.\pi m/n\right)-1}\\
&= \frac{1}{\mathbf{i}}\frac{\exp\left(\mathbf{i}.\pi m/(2n)\right)+\exp\left(-\mathbf{i}.\pi m/(2n)\right)}{\exp\left(\mathbf{i}.\pi m/(2n)\right)-\exp\left(-\mathbf{i}.\pi m/(2n)\right)}\left(\cos\left(\pi m\right)-1\right)\\
&= -\frac{\cos\left(\pi m/(2n)\right)}{\sin\left(\pi m/(2n)\right)}\left(\cos\left(\pi m\right)-1\right)
\end{aligned}
$$

Thus, for $j_1 \neq j_2$,

$$
\begin{aligned}
&\sum_{k=1}^{n} p_{j_1}(x_{n,k}|w_{\mathcal{C}[-1,1]})p_{j_2}(x_{n,k}|w_{\mathcal{C}[-1,1]})\\
&= \frac{1}{2}\cos\left(0.5\pi(j_1+j_2)/n\right)\left(\cos\left(\pi(j_1+j_2)\right)-1\right)\\
&-\frac{1}{2}\cos\left(\pi(j_1+j_2)/(2n)\right)\left(\cos\left(\pi(j_1+j_2)\right)-1\right)\\
&+\frac{1}{2}\cos\left(0.5\pi(j_1-j_2)/n\right)\left(\cos\left(\pi(j_1-j_2)\right)-1\right)\\
&-\frac{1}{2}\cos\left(\pi(j_1-j_2)/(2n)\right)\left(\cos\left(\pi(j_1-j_2)\right)-1\right)\\
&= 0
\end{aligned}
$$

whereas for $j_1 = j_2 = j > 0$,

$$
\begin{aligned}
&\sum_{k=1}^{n} p_{j}(x_{n,k}|w_{\mathcal{C}[-1,1]})p_{j}(x_{n,k}|w_{\mathcal{C}[-1,1]})\\
&= \frac{1}{2}\cos\left(\pi.j/n\right)\sum_{k=1}^{n}\cos\left(2\pi.j.k/n\right) + \frac{1}{2}\sin\left(\pi.j/n\right)\sum_{k=1}^{n}\sin\left(2\pi.j.k/n\right)\\
&+\frac{1}{2}n = \frac{1}{2}n
\end{aligned}
$$

The case $j_1 = j_2 = 0$ is trivial.

Chapter 5

Trigonometric series

5.1 Cosine series representation

Note that the distribution function $W_{\mathcal{C}[0,1]}(u)$ defined by (4.22) has inverse

$$W_{\mathcal{C}[0,1]}^{-1}(u) = (1 + \cos(\pi(1-u)))/2. \tag{5.1}$$

It is now easy to verify from Corollary 4.2, (5.1) and (4.24) that every function $f \in L^2(0,1)$ can be written as

$$\begin{aligned} f(u) &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2} \cos\left(k. \arccos(2W_c^{-1}(u) - 1)\right) \\ &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2} \cos(k\pi(1-u)) \\ &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k (-1)^k \sqrt{2} \cos(k\pi u) \\ &= \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos(k\pi u) \end{aligned}$$

where

$$\begin{aligned} \alpha_k &= \gamma_k (-1)^k = (-1)^k \int_0^1 f(u) p_k\left(\frac{1}{2}(1 + \cos(\pi(1-u)))\middle| w_{\mathcal{C}[0,1]}\right) du \\ &= \begin{cases} \int_0^1 f(u)\, du & \text{if} \quad k = 0, \\ \int_0^1 f(u) \sqrt{2} \cos(k\pi u)\, du & \text{if} \quad k \in \mathbb{N}. \end{cases} \end{aligned}$$

Consequently,

Theorem 5.1. *The functions*

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

form a complete orthonormal sequence in $L^2(0,1)$. *Thus, given a function* $f \in L^2(0,1)$, *let*

$$f_n(u) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi u)$$

where $\alpha_k = \int_0^1 f(u)\kappa_k(u)\,du$. *Then* $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$ *and*

$$\lim_{n\to\infty} \int_0^1 (f(u) - f_n(u))^2\,du = \lim_{n\to\infty} \sum_{k=n+1}^{\infty} \alpha_k^2 = 0.$$

Consequently, similar to Theorem 4.1, f can be written as

$$f(u) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2}\cos(k\pi u) \ \ a.e. \ on \ (0,1). \tag{5.2}$$

5.2 Fourier analysis

Consider the following sequence of functions on $[-1,1]$:

$$\begin{aligned} \varphi_0(x) &= 1 \\ \varphi_{2k-1}(x) &= \sqrt{2}\sin(k\pi x), \ \varphi_{2k}(x) = \sqrt{2}\cos(k\pi x), \ k \in \mathbb{N}. \end{aligned} \tag{5.3}$$

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 weight function $w(x) = \frac{1}{2}I(|x| \leq 1)$, i.e.,

$$\frac{1}{2}\int_{-1}^{1} \varphi_m(x)\varphi_k(x)\,dx = I(m = k)$$

It is a classical Fourier analysis result that

Theorem 5.2. *The Fourier series* $\{\varphi_n\}_{n=0}^{\infty}$ *is complete in* $L^2(-1,1)$.

The "official" proof of this result is long and tedious. See for example Young (1988). However, using Theorem 5.1 this result can be proved somewhat easier, as follows.

We need to show that for an arbitrary function $g \in L^2(-1,1)$,

$$\lim_{n\to\infty} \frac{1}{2}\int_{-1}^{1} (g(x) - g_n(x))^2\, dx, \tag{5.4}$$

where

$$g_n(x) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi x) + \sum_{k=1}^{n} \beta_k \sqrt{2}\sin(k\pi x) \tag{5.5}$$

with Fourier coefficients

$$\begin{aligned} \alpha_0 &= \frac{1}{2}\int_{-1}^{1} g(x)\, dx \\ \alpha_k &= \frac{1}{2}\int_{-1}^{1} \sqrt{2}\cos(k\pi x)\, g(x)\, dx \\ \beta_k &= \frac{1}{2}\int_{-1}^{1} \sqrt{2}\sin(k\pi x)\, g(x)\, dx. \end{aligned}$$

Let $x = 2u - 1$ for $u \in [0,1]$, and denote

$$f(u) = g(2u-1),\ f_n(u) = g_n(2u-1)$$

Then it follows from the well-known sine-cosine equalities that

$$\begin{aligned} \alpha_0 &= \int_0^1 g(2u-1)\, du = \int_0^1 f(u)du \\ \alpha_k &= \int_0^1 \sqrt{2}\cos(k\pi(2u-1))\, g(2u-1)\, du \\ &= (-)^k \int_0^1 \sqrt{2}\cos(2k\pi u)\, f(u)du \\ \beta_k &= \int_0^1 \sqrt{2}\sin(k\pi(2u-1))\, g(2u-1)\, du \\ &= (-)^k \int_0^1 \sqrt{2}\sin(2k\pi u)\, f(u)du \end{aligned}$$

and

$$\begin{aligned} f_n(u) &= g_n(2u-1) \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi(2u-1)) + \sum_{k=1}^{n} \beta_k \sqrt{2}\sin(k\pi(2u-1)) \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k (-)^k \sqrt{2}\cos(2k\pi u) + \sum_{k=1}^{n} \beta_k (-)^k \sqrt{2}\sin(2k\pi u) \end{aligned}$$

Thus, (5.4) is true if and only

$$\lim_{n\to\infty} \int_0^1 (f(u) - f_n(u))^2 \, du = 0.$$

Theorem 5.2 follows now from the following result, which will be proved in the appendix to this chapter.

Theorem 5.3. *The functions* $\overline{\varphi}_0(u) = 1$, $\overline{\varphi}_k(u) = \sqrt{2}\sin(2k\pi u)$ *if* $k \geq 1$ *is odd,* $\overline{\varphi}_k(u) = \sqrt{2}\cos(2k\pi u)$ *if* $k \geq 2$ *is even, form a complete orthonormal sequence in* $L^2(0,1)$.

Although Theorem 5.1 was used to prove Theorem 5.2, Theorem 5.2 can also be proved independently. See for example Young (1988). Then Theorem 5.1 becomes a corollary of Theorem 5.2, as follows.

Let $f(u) \in L^2(0,1)$ be arbitrary, and let $g(x) = f(|x|)$. Then $g(x) \in L^2(-1,1)$, with Fourier coefficients

$$\begin{aligned} \alpha_0 &= \frac{1}{2}\int_{-1}^{1} f(|x|)\, dx = \int_0^1 f(u)\, du \\ \alpha_k &= \frac{1}{2}\int_{-1}^{1} \sqrt{2}\cos(k\pi x) f(|x|)\, dx = \int_0^1 \sqrt{2}\cos(k\pi u) f(u)\, du \\ \beta_k &= \frac{1}{2}\int_{-1}^{1} \sqrt{2}\sin(k\pi x) f(|x|)\, dx = 0 \end{aligned}$$

Hence it follows from Theorem 5.2 that

$$\lim_{n\to\infty} \int_0^1 \left(f(u) - \alpha_0 - \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi u) \right)^2 du \qquad (5.6)$$

$$= \frac{1}{2} \lim_{n\to\infty} \int_{-1}^{1} \left(f\left(|x|\right) - \alpha_0 - \sum_{k=1}^{n} \alpha_k \sqrt{2} \cos\left(k\pi x\right) \right)^2 dx$$
$$= 0$$

Similar to the proof of Theorem 4.1 is follows now from (5.6) that

$$f\left(u\right) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos\left(k\pi u\right) \text{ a.e. on } (0,1),$$

where $\alpha_0 = \int_0^1 f\left(u\right) du$ and $\alpha_k = \int_0^1 \sqrt{2} \cos\left(k\pi u\right) f\left(u\right) du$ for $k \geq 1$, which is just the result in Theorem 5.1.

5.3 Sine series representation

Let $f(x)$ be a square integrable function on $[-1,1]$ such that $f(x) = -f(-x)$, with a possible discontinuity at $x = 0$. Then

$$\begin{aligned}
\beta_k &= \frac{1}{2} \int_{-1}^{1} f(x) \sqrt{2} \sin\left(k\pi x\right) du = \int_0^1 f(u) \sqrt{2} \sin\left(k\pi u\right) du \\
0 &= \frac{1}{2} \int_{-1}^{1} f(x) \sqrt{2} \cos\left(k\pi x\right) dx \\
0 &= \frac{1}{2} \int_{-1}^{1} f(x) dx = 0
\end{aligned}$$

Hence by Theorem 5.2, $\lim_{n\to\infty} \frac{1}{2} \int_{-1}^{1} \left(f(x) - \sum_{k=1}^{n} \beta_k \sqrt{2} \sin\left(k\pi x\right) \right)^2 dx = 0$, which by the condition $f(x) = -f(-x)$ implies

$$\lim_{n\to\infty} \int_0^1 \left(f(u) - f_n(u)\right)^2 du = 0,$$

where

$$f_n(u) = \sum_{k=1}^{n} \beta_k \sqrt{2} \sin\left(k\pi u\right)$$

Moreover, it is easy to verify that

$$\int_0^1 \sqrt{2} \sin\left(k\pi u\right) \sqrt{2} \sin\left(m\pi u\right) du = I(k = m).$$

Thus, we have the following corollary of Theorem 5.2.

Theorem 5.4. *The sine series* $\{\sqrt{2}\sin(k\pi u)\}_{k=1}^{\infty}$ *is a complete orthonormal sequence in* $L^2(0,1)$. *Consequently, any function* $f \in L^2(0,1)$ *can be written as*

$$f(u) = \sum_{k=1}^{\infty} \beta_k \sqrt{2}\sin(k\pi u) \text{ a.e. on } (0,1),$$

where $\beta_k = \int_0^1 f(u)\sqrt{2}\sin(k\pi u)\,du$.

Note however that $f_n(u)$ will be a poor approximation of $f(u)$ for u close to zero or one because $f_n(0) = f_n(1) = 0$ whereas $f(0)$ and $f(1)$ may be nonzero. The reason is that in general $\lim_{u\to u_0}\lim_{n\to\infty} f_n(u) \neq \lim_{n\to\infty}\lim_{u\to u_0} f_n(u)$.

5.4 How well does the cosine series fit?

5.4.1 Exact Fourier coefficients

To check how well the cosine series fit, consider the function $f(u) = u(4-3u)$ on $[0,1]$. Note that this is a density function. For this function we can derive the Fourier coefficients involved analytically, as

$$\begin{aligned} \alpha_0 &= \int_0^1 f(u)du = 1, \\ \alpha_k &= \int_0^1 f(u)\sqrt{2}\cos(k\pi u)du = -2\sqrt{2}(k\pi)^{-2}\left((-1)^k + 1\right) \end{aligned}$$

This way of approximating densities directly by a series expansion has been advocated by Kronmal and Tarter (1968). However, a potential problem with this approach is that in general there is no guarantee that $f_n(u) \geq 0$.

In the following figures the function $f(u) = u(4-3u)$ is compared with its SNP approximation $f_n(u) = 1 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi u)$ (dotted curve) for $n = 4, 8, 12$.

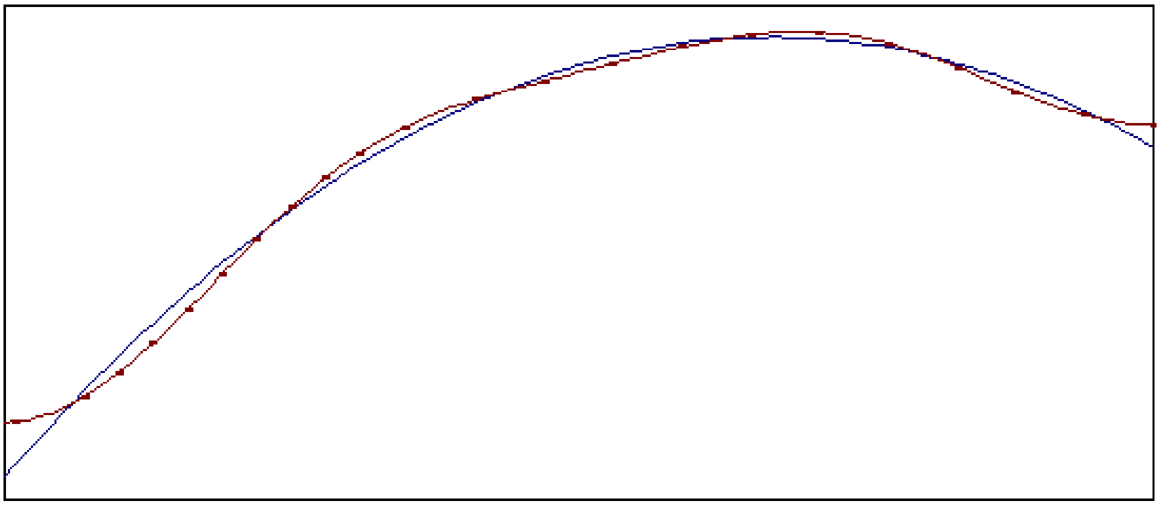

Figure 5.1: $f(u) = u(4 - 3u)$ compared with $f_n(u)$ for $n = 4$

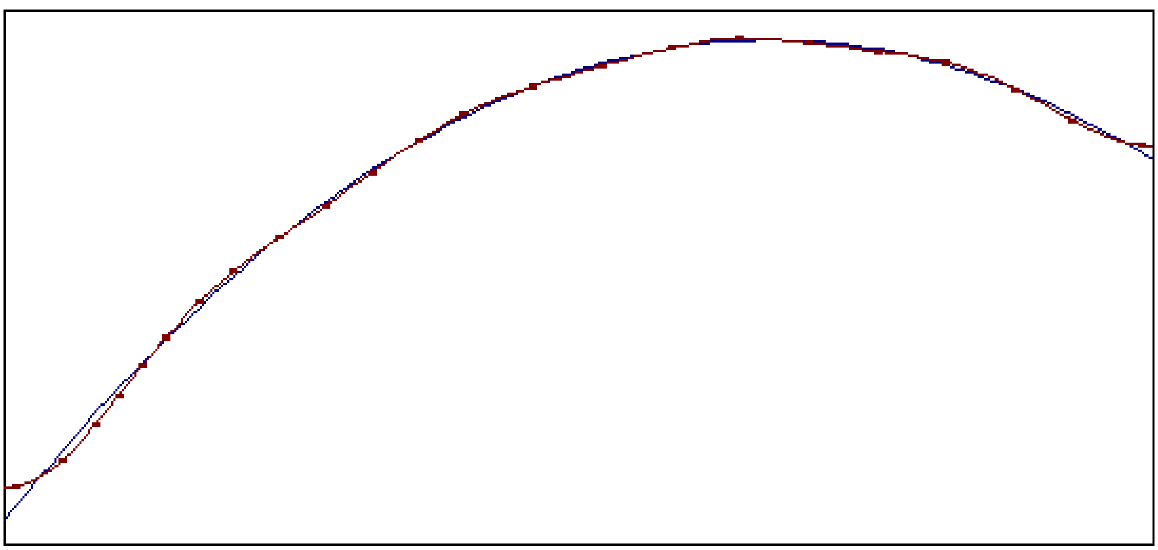

Figure 5.2: $f(u) = u(4 - 3u)$ compared with $f_n(u)$ for $n = 8$

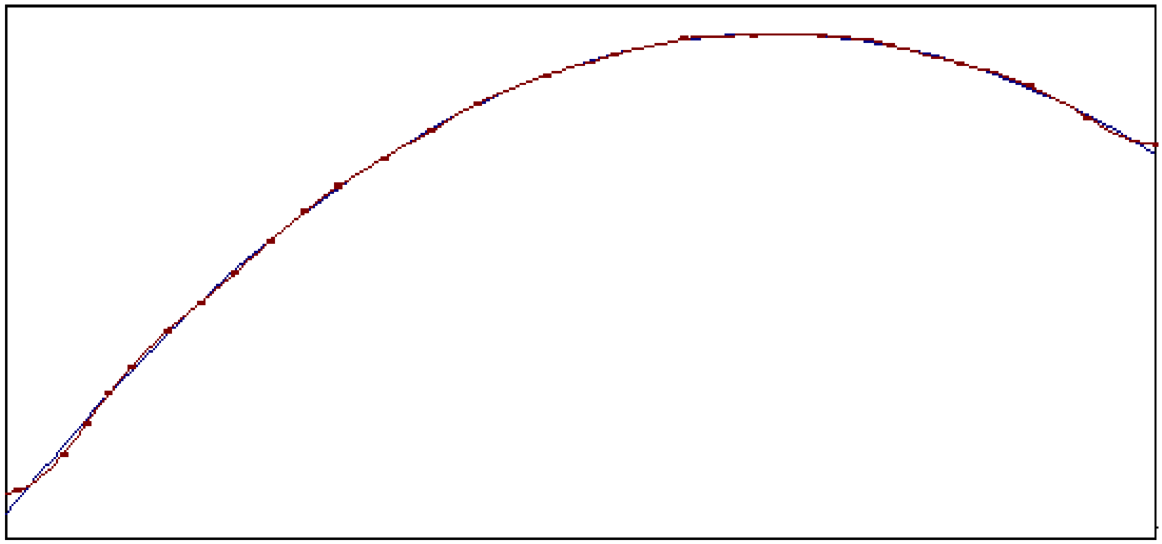

Figure 5.3: $f(u) = u(4-3u)$ compared with $f_n(u)$ for $n = 12$

We see that $f_n(u)$ approximates $f(u)$ quite well, even for $n = 4$, except for the tails of $f_n(u)$ in the latter case. The reason is that $f_n'(u) = -\sum_{k=1}^{n} \alpha_k k\pi\sqrt{2}\sin(k\pi u)$, so that $f_n'(0) = f_n'(1) = 0$. As expected, the tail fit becomes better for larger truncation orders n.

5.4.2 Bivariate SNP regression

Let $(Y, X) \in \mathbb{R}^2$ be a pair of absolutely continuous random variables satisfying

$$E[Y^2] < \infty,\ E[X^2] < \infty. \tag{5.7}$$

We can always write

$$E[Y|X] = f(X) = \alpha + \beta X + X^2 r(X), \tag{5.8}$$

where $\alpha+\beta X$ is the linear projection of Y on 1 and X, with residual $X^2 r(X)$. Moreover, given an absolutely continuous distribution function $G(x)$ with density $g(x) > 0$ on $\mathbb{R}$ and inverse $G^{-1}(u)$, $u \in [0,1]$, we can write

$$r(x) = \varphi(G(x)) \tag{5.9}$$

where

$$\varphi(u) = r(G^{-1}(u)) \tag{5.10}$$

Now let us assume that

$$\int_0^1 \varphi(u)^2 du = \int_{-\infty}^{\infty} r(x)^2 g(x) dx < \infty \tag{5.11}$$

so that $\varphi \in L^2(0,1)$. Then by Theorem 5.1, φ has the series expansion

$$\varphi(u) = \gamma + \sum_{k=1}^{\infty} \delta_k \sqrt{2} \cos(k\pi u) \text{ a.e. on } [0,1],$$

where

$$\gamma = \int_0^1 \varphi(u) du, \ \delta_k = \int_0^1 \sqrt{2} \cos(k\pi u) \varphi(u) du.$$

Consequently,

$$f(X) = E[Y|X] = \alpha + \beta X + \gamma X^2 + X^2 \sum_{k=1}^{\infty} \delta_k \sqrt{2} \cos(k\pi G(X)) \text{ a.s.}$$

Next, let

$$f_n(X) = \alpha + \beta X + \gamma X^2 + X^2 \sum_{k=1}^{n} \delta_k \sqrt{2} \cos(k\pi G(X)),$$

and denote $r_n(x) = \sum_{k=1}^n \delta_k \sqrt{2} \cos(k\pi G(X))$. Since by Theorem 5.1,

$$\lim_{n\to\infty} r_n(x) = r(x) \text{ a.e.},$$

it follows that

$$\lim_{n\to\infty} f_n(X) = f(X) \text{ a.s.}$$

In principle we could specify $f(X)$ directly as $f(X) = \varphi(G(x))$, but if $f(x)$ is linear then we need the full series expansion of φ to fit $f(x) = \alpha + \beta X$, whereas in the case (5.8) the linear regression model corresponds to $r(x) \equiv 0$.

A convenient choice for G is the logistic distribution function

$$G(x) = (1 + \exp(-x))^{-1},$$

which has density $g(x) = G(x)(1 - G(x))$ and inverse $G^{-1}(u) = \ln(u/(1-u))$. Since all the moments of the Logistic distribution are finite, the condition (5.11) allows $r(x)$ to be a polynomial of any order.

To check how well $f_n(X)$ fits, let $Y = f(X) + U$, where X and U independent standard normally distributed, and

$$f(x) = (|x| - 1/4)^3 \left(I\left(x > 1/4\right) - I\left(x < -1/4\right)\right).$$

The reason for this choice of $f(x)$ is to check whether the cosine series expansion is able to capture the horizontal part of $f(x)$ for $|x| \leq 1/4$.

The following three figures compare $f(x)$ with the SNP-OLS estimates $\widehat{f}_n(x)$ of $f_n(x)$ for $n = 4, 8, 12$ and $x \in [-2, 2]$, with G the Logistic distribution function, on the basis of a random sample of size 500 from (Y, X).

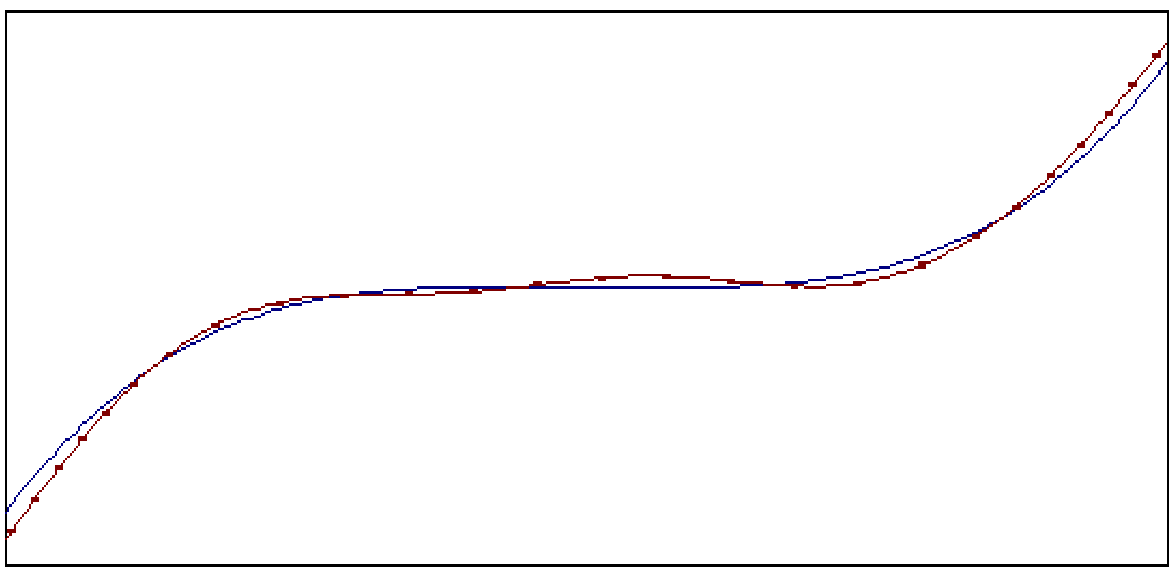

Figure 5.4: $f(x)$ compared with its SNP-OLS estimate $\widehat{f}_4(x)$ on $[-2, 2]$

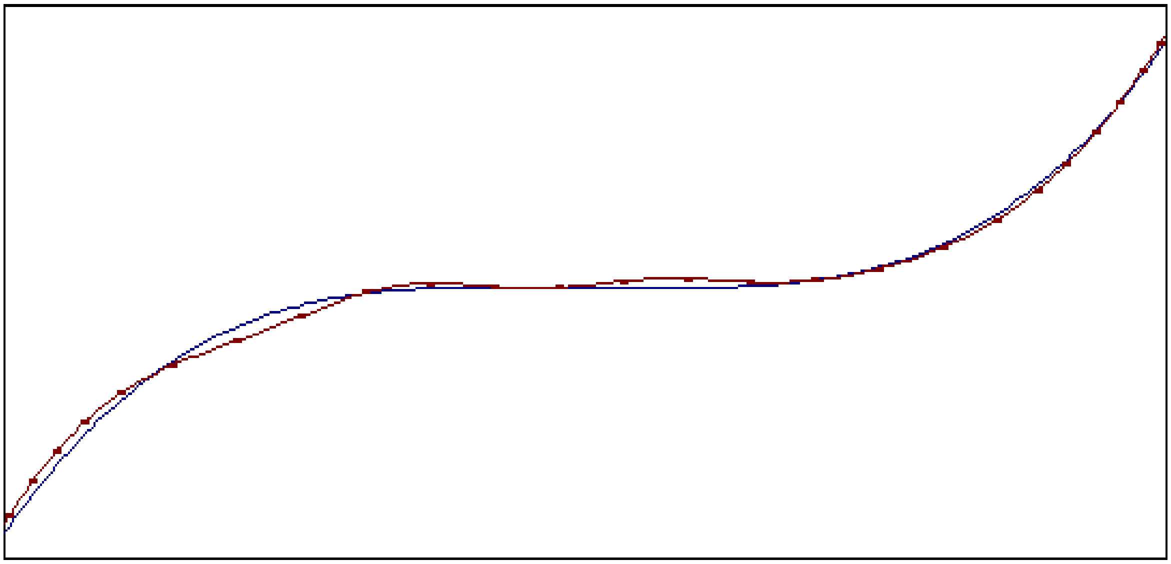

Figure 5.5: $f(x)$ compared with its SNP-OLS estimate $\widehat{f}_8(x)$ on $[-2, 2]$

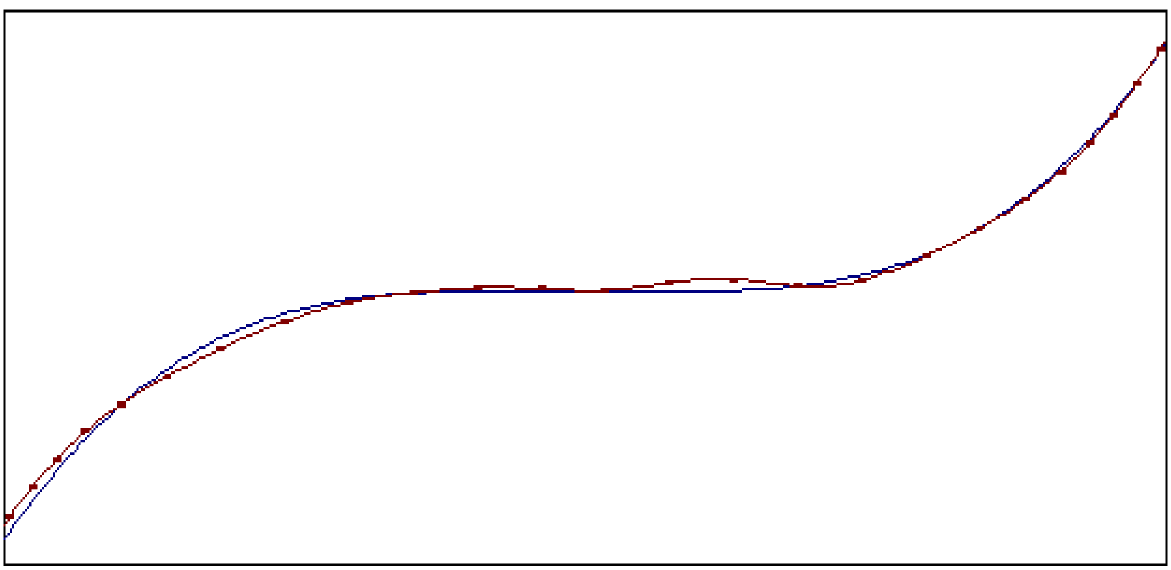

Figure 5.6: $f(x)$ compared with its SNP-OLS estimate $\widehat{f}_{12}(x)$ on $[-2, 2]$

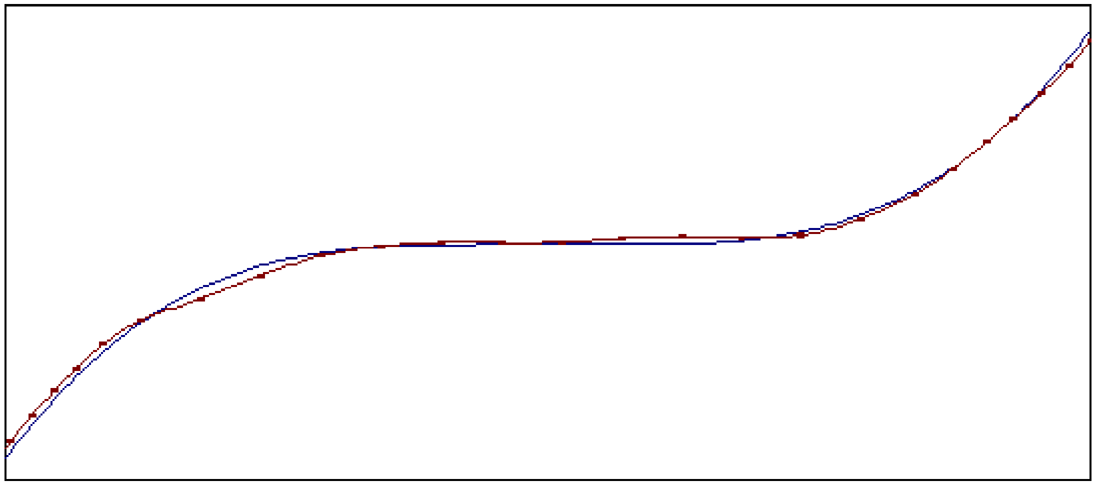

Figure 5.7: Comparison of $f(x)$ with its nonparametric kernel regression estimate $\widetilde{f}(x)$ on $[-2, 2]$

As a comparison I have also estimated $f(x)$ by nonparametric kernel regression, similar to Bierens and Pott-Buter (1990), with standard normal kernel and bandwidth constant determined by in-sample leaving-one-out cross-validation over the interval $[0.1, 2]$. The result for $x \in [-2, 2]$ is displayed in Figure 5.7.

As to the SNP results, note the slight wiggle of $\widehat{f}_n(x)$ in the flat area $|x| < 1/4$, whereas the nonparametric kernel regression estimator $\widetilde{f}(x)$ is smoother in this area. However, in view of the fact that this flat part of $f(x)$ has been approximated via a linear combination of cosine functions the SNP approach works better than I expected.

5.5 Appendix: Proof of Theorem 5.3

The orthonormality of the sequence $\{\overline{\varphi}_n\}_{n=0}^{\infty}$ is easy to verify. The completeness proof employs the following steps.

Step 1. Let $C_0[0,1]$ be the space of continuous functions $f(u)$ on $[0,1]$ satisfying $\int_0^1 f(u)du = 0$, endowed with the $L^2(0,1)$ topology, and let $C_{0,1}[0,1]$ be the space of continuously differentiable functions $F(u)$ on $[0,1]$ satisfying $F(0) = F(1) = 0$, also endowed with the $L^2(0,1)$ topology. Note that the functions in $C_{0,1}[0,1]$ take the form $F(u) = \int_0^u f(x)\,dx$ with $f(u) = F'(u)$. It will be shown that $C_{0,1}[0,1] \subset \overline{\text{span}}(\{\varphi_n\}_{n=0}^\infty)$.

Step 2. It will be shown that $C_0[0,1]$ is the closure of $C_{0,1}[0,1]$, hence $C_0[0,1] \subset \overline{\text{span}}(\{\varphi_n\}_{n=0}^\infty)$. It follows then trivially that the space $C[0,1]$ of continuous functions on $[0,1]$ is contained in $\overline{\text{span}}(\{\varphi_n\}_{n=0}^\infty)$.

Step 3. Finally, it will be shown that every function in $L^2(0,1)$ can be written as a limit of a sequence of continuous functions, hence $L^2(0,1)$ is the closure of $C[0,1]$, so that $L^2(0,1) = \overline{\text{span}}(\{\varphi_n\}_{n=0}^\infty)$.

Proof of Step 1

Let $f_n(u)$ and $f(u)$ be the same as in Theorem 5.1, except that due to the condition $\int_0^1 f(u)du = 0$, $\alpha_0 = 0$, and let $F_n(u) = \int_0^u f_n(x)\,dx$. Then

$$\begin{aligned} F_n(u) &= \sum_{k=1}^{n} \frac{\alpha_k}{k\pi}\sqrt{2}\sin(k\pi u) \\ &= \sum_{k=1}^{[(n+1)/2]} \frac{\alpha_{2k-1}}{(2k-1)\pi}\sqrt{2}\sin((2k-1)\pi u) + \sum_{k=1}^{[n/2]} \frac{\alpha_{2k}}{2k\pi}\sqrt{2}\sin(2k\pi u) \end{aligned}$$

and

$$\begin{aligned} \sup_{0\le u\le 1} |F(u) - F_n(u)| &\le \int_0^1 |f(x) - f_n(x)|\,dx \\ &\le \sqrt{\int_0^1 (f(x) - f_n(x))^2\,dx} = o(1) \qquad (5.12) \end{aligned}$$

Next, observe that

$$\int_0^1 \sqrt{2}\sin((2k-1)\pi u)\,du = \frac{-2\sqrt{2}}{(2k-1)\pi} = \gamma_{0,k}$$

$$\int_0^1 \sqrt{2}\sin((2k-1)\pi u)\sqrt{2}\cos((2m-1)\pi u)\,du = 0$$

$$\begin{aligned} &\int_0^1 \sqrt{2}\sin((2k-1)\pi u)\sqrt{2}\cos(2m\pi u)\,du \\ &\quad = \frac{-2}{(2(k+m)-1)\pi} + \frac{-2}{(2(k-m)-1)\pi} \\ &\quad = -\frac{2}{\pi}\frac{4k-2}{(2(k+m)-1)(2(k-m)-1)} \end{aligned}$$

$$= -\frac{2}{\pi}\frac{k-1/2}{(k-1/2)^2 - m^2} = \gamma_{m,k}$$

Hence

$$\sqrt{2}\sin\left((2k-1)\pi u\right) = \gamma_{0,k} + \sum_{m=1}^{\infty}\gamma_{m,k}\sqrt{2}\cos\left(2m\pi u\right)$$

a.e. on $[0,1]$. Now let

$$\begin{aligned}
\widetilde{F}_n(u) &= \sum_{k=1}^{[(n+1)/2]}\frac{\alpha_{2k-1}}{(2k-1)\pi}\gamma_{0,k} + \sum_{k=1}^{[n/2]}\frac{\alpha_{2k}}{2k\pi}\sqrt{2}\sin\left(2k\pi u\right) \\
&\quad + \sum_{m=1}^{N}\left(\sum_{k=1}^{[(n+1)/2]}\frac{\alpha_{2k-1}}{(2k-1)\pi}\gamma_{m,k}\right)\sqrt{2}\cos\left(2m\pi u\right) \\
&= -2\sqrt{2}\sum_{k=1}^{[(n+1)/2]}\frac{\alpha_{2k-1}}{(2k-1)^2\pi^2} + \sum_{k=1}^{[n/2]}\frac{\alpha_{2k}}{2k\pi}\sqrt{2}\sin\left(2k\pi u\right) \\
&\quad - \frac{1}{\pi^2}\sum_{m=1}^{N}\left(\sum_{k=1}^{[(n+1)/2]}\frac{\alpha_{2k-1}}{(k-1/2)^2 - m^2}\right)\sqrt{2}\cos\left(2m\pi u\right)
\end{aligned}$$

where $N \geq [(n+1)/2]$. Then

$$\begin{aligned}
\int_0^1\left(\widetilde{F}_n(u) - F_n(u)\right)^2 du &= \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\sum_{k=1}^{[(n+1)/2]}\frac{\alpha_{2k-1}}{m^2 - (k-1/2)^2}\right)^2 \\
&\leq \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\sum_{k=1}^{[(n+1)/2]}\frac{|\alpha_{2k-1}|}{m^2 - (k-1/2)^2}\right)^2 \\
&\leq \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\frac{\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{m^2 - ([n/2])^2}\right)^2 \\
&= \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\frac{\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{(m-[n/2])(m+[n/2])}\right)^2 \\
&\leq \frac{1}{4\pi^4}\sum_{m=N+1}^{\infty}\left(\frac{\frac{1}{[n/2]}\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{m-[n/2]}\right)^2
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{4\pi^4}\sum_{m=[n/2]+1}^{\infty}\left(\frac{\frac{1}{[n/2]}\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{m-[n/2]}\right)^2 \\
&= \frac{1}{4\pi^4}\left(\sum_{m=1}^{\infty}\frac{1}{m^2}\right)\left(\frac{1}{[n/2]}\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|\right)^2 \\
&\leq \frac{1}{4\pi^4}\left(\sum_{m=1}^{\infty}\frac{1}{m^2}\right)\frac{1}{[n/2]}\sum_{k=1}^{\infty}\alpha_{2k-1}^2 \\
&= O\left(1/n\right)) \qquad (5.13)
\end{aligned}
$$

Hence by (5.12) and (5.13),

$$\lim_{n\to\infty}\int_0^1\left(\widetilde{F}_n(u)-F(u)\right)^2du=0$$

Since $\widetilde{F}_n \in \operatorname{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$ it follows that $F \in \operatorname{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$, hence $C_{0,1}[0,1] \subset \operatorname{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

Proof of Step 2

Choose an arbitrary function $f \in C_0[0,1]$, and extend $f(x)$ for $x > 1$ as $f(x) = f(1)$. Let $F(u) = \int_0^u f(x)\,dx$ and

$$f_n(u) = \frac{(F(u+n^{-1})-F(u))}{n^{-1}} = \frac{1}{n}\int_u^{u+1/n} f(x)dx$$

Then by continuity

$$\lim_{n\to\infty}|f_n(u)-f(u)| \leq \lim_{n\to\infty}\sup_{u\leq x\leq u+1/n}|f(x)-f(u)|=0$$

pointwise in $u \in [0,1]$. Moreover,

$$\sup_{0\leq u\leq 1}|f_n(u)-f(u)| \leq 2\sup_{0\leq u\leq 1}|f(u)|<\infty$$

Therefore if follows by bounded convergence that

$$\lim_{n\to\infty}\int_0^1(f_n(u)-f(u))^2\,du=0$$

Since $f_n \in C_{0,1}[0,1] \subset \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$ it follows now that $C_0[0,1] \subset \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

Because the functions in $C[0,1]$ differ from the functions in $C_0[0,1]$ by constants only, it follows that $C[0,1] \subset \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

Proof of Step 3

Let B be an arbitrary Borel subset of $[0,1]$ and let

$$f_n(u) = \exp\left(-n^{-1}\inf_{x\in\overline{B}}|x-u|\right) - \exp\left(-n^{-1}\inf_{x\in\overline{B}\backslash B}|x-u|\right),$$

where $\overline{B}$ is the closure of B. This function is continuous on $[0,1]$. To see this, note that for $u_1, u_2 \in [0,1]$,

$$\begin{aligned} \inf_{x\in B}|x-u_1| &\leq |u_2-u_1| + \inf_{x\in B}|x-u_2| \\ \inf_{x\in\overline{B}\backslash B}|x-u_2| &\leq |u_2-u_1| + \inf_{x\in\overline{B}\backslash B}|x-u_1| \end{aligned}$$

hence

$$\left|\inf_{x\in\overline{B}}|x-u_2| - \inf_{x\in\overline{B}}|x-u_1|\right| \leq |u_2-u_1|$$

and similarly,

$$\left|\inf_{x\in\overline{B}\backslash B}|x-u_2| - \inf_{x\in\overline{B}\backslash B}|x-u_1|\right| \leq |u_2-u_1|$$

For $u \in B$, $\inf_{x\in\overline{B}}|x-u| = 0$ and $\inf_{x\in\overline{B}\backslash B}|x-u| > 0$, hence $\lim_{n\to\infty} f_n(u) = 1$. For $u \in B\backslash\overline{B}$, $\inf_{x\in\overline{B}}|x-u| = 0$ and $\inf_{x\in\overline{B}\backslash B}|x-u| = 0$, hence $f_n(u) = 0$, and for $u \in [0,1]\backslash\overline{B}$, $\inf_{x\in\overline{B}}|x-u| > 0$ and $\inf_{x\in\overline{B}\backslash B}|x-u| > 0$, hence $\lim_{n\to\infty} f_n(u) = 0$. Thus

$$\lim_{n\to\infty} f_n(u) = I(x \in B).$$

Since $f_n(u) \in C[0,1] \subset \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$ it follows now that for arbitrary Borel sets B, $I(x\in B) \in \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$ and so are all simple functions on $[0,1]$. Because functions are Borel measurable if and only if they are limits of sequences of simple functions, it follows that $L^2(0,1) = \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

■

Chapter 6

Density and distribution functions

6.1 Density functions on the unit interval

It follows from Theorem 5.1 that for any density function $h(u)$ on $[0,1]$ there exists a sequence $\{\alpha_k\}_{k=0}^{\infty}$ satisfying $\sum_{k=0}^{\infty}\alpha_k^2 = 1$ such that

$$h\left(u\right) = \left(\alpha_0 + \sum_{k=1}^{\infty}\alpha_k\sqrt{2}\cos\left(k\pi u\right)\right)^2 \text{ a.e. on } (0,1). \tag{6.1}$$

The square guarantees that $h\left(u\right) \geq 0$. Gallant and Nychka (1987) proposed a similar series expansion on the basis of Hermite polynomials.

Note that the α_k's in (6.1) are no longer unique. For example, we can always write $h\left(u\right) = f_B\left(u\right)^2$, where for an arbitrary Borel set B in $[0,1]$,

$$f_B\left(u\right) = \left(I\left(u \in B\right) - I\left(u \notin B\right)\right)\sqrt{h(u)}. \tag{6.2}$$

Then the α_k's in (6.1) take the form

$$\alpha_k = \int_B \sqrt{h(u)}\kappa_k\left(u\right)du - \int_{[0,1\backslash B}\sqrt{h(u)}\kappa_k\left(u\right)du$$

In particular, we may choose for α_0 any

$$\alpha_0 \in \left[-\int_0^1\sqrt{h(u)}du, \int_0^1\sqrt{h(u)}du\right]. \tag{6.3}$$

If we choose $\alpha_0 \in \left(0, \int_0^1 \sqrt{h(u)}du\right]$ then we can reparametrize the Fourier coefficients α_k as

$$\begin{aligned}\alpha_0 &= \frac{1}{\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2}} \\ \alpha_k &= \frac{\delta_k}{\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2}},\ k \in \mathbb{N},\end{aligned}$$

where $\sum_{m=0}^{\infty}\delta_m^2 < \infty$. Hence,

Theorem 6.1. *For any density function $h(u)$ on $[0,1]$ there exist possibly uncountable many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=0}^{\infty}\delta_m^2 < \infty$ such that*

$$h(u) = \frac{\left(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos(k\pi u)\right)^2}{1+\sum_{m=1}^{\infty}\delta_m^2} \ a.e.\ on\ (0,1). \tag{6.4}$$

In particular, (6.4) *holds for all sequences δ_k of the form*

$$\delta_k = \frac{\int_0^1 \left(I(u \in B) - I(u \notin B)\right)\sqrt{h(u)}\sqrt{2}\cos(k\pi u)\,du}{\int_0^1 \left(I(u \in B) - I(u \notin B)\right)\sqrt{h(u)}du}, \tag{6.5}$$

where B is any Borel set in $[0,1]$ satisfying

$$\int_0^1 \left(I(u \in B) - I(u \notin B)\right)\sqrt{h(u)}du > 0.$$

Moreover, the corresponding SNP densities

$$h_n(u) = \frac{\left(1+\sqrt{2}\sum_{k=1}^{n}\delta_k\cos(k\pi u)\right)^2}{1+\sum_{m=1}^{n}\delta_m^2} \tag{6.6}$$

satisfy

$$\int_0^1 |h(u) - h_n(u)|\,du \le \sqrt{5\sum_{k=n+1}^{\infty}\delta_k^2} \to 0 \tag{6.7}$$

Furthermore, the corresponding SNP distribution functions have the closed form expressions

$$H_n(u) = u$$

$$+\frac{1}{1+\sum_{m=1}^{n}\delta_m^2}\left[2\sqrt{2}\sum_{k=1}^{n}\delta_k\frac{\sin(k\pi u)}{k\pi}+\sum_{m=1}^{n}\delta_m^2\frac{\sin(2m\pi u)}{2m\pi}\right. \tag{6.8}$$
$$\left.+2\sum_{k=2}^{n}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin((k+m)\pi u)}{(k+m)\pi}+2\sum_{k=2}^{n}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin((k-m)\pi u)}{(k-m)\pi}\right],$$

and satisfy

$$\sup_{0\leq u\leq 1}|H(u)-H_n(u)|\leq\sqrt{5\sum_{k=n+1}^{\infty}\delta_k^2}\to 0. \tag{6.9}$$

6.2 Uniqueness of the series representation

The density $h(u)$ in Theorem 6.1 can be written as $h(u)=\eta(u)^2/\int_0^1\eta(v)^2dv$, where

$$\eta(u)=1+\sum_{m=1}^{\infty}\delta_m\sqrt{2}\cos(m\pi u)\ \text{ a.e. on } (0,1). \tag{6.10}$$

Moreover, recall that in general,

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

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

$$\eta(u)=(I(u\in B)-I(u\in[0,1]\backslash B))\sqrt{h(u)}\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2} \tag{6.11}$$

Similarly, given this Borel set B and the corresponding δ_m's, the SNP density (6.6) can be written as $h_n(u)=\eta_n(u)^2/\int_0^1\eta_n(v)^2dv$, where

$$\eta_n(u)\ =\ 1+\sum_{m=1}^{n}\delta_m\sqrt{2}\cos(m\pi u)$$

$$= \quad (I(u \in B) - I(u \in [0,1]\backslash B))\sqrt{h_n(u)}\sqrt{1+\sum_{m=1}^{n}\delta_m^2} \quad (6.12)$$

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 $u_0 \in (0,1)\backslash S$, $\lim_{n\to\infty} h_n(u_0) = h(u_0) > 0$, hence for sufficient large n, $h_n(u_0) > 0$. 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

$$I(u \in B) - I(u \in [0,1]\backslash B) = \frac{\eta_n(u)}{\sqrt{h_n(u)}\sqrt{1+\sum_{m=1}^{n}\delta_m^2}} \quad (6.13)$$

is continuous on $(u_0-\varepsilon_n(u_0), u_0+\varepsilon_n(u_0))\cap(0,1)$. Substituting (6.13) in (6.11) 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 (6.10), $\int_0^1 \eta(u)du = 1$. Thus, $\eta(u) > 0$ on $(0,1)$, so that by (6.11), $I(u \in B) - I(u \in [0,1]\backslash B) = 1$ on $(0,1)$.

Consequently,

Theorem 6.2. *For every continuous and positive valued density $h(u)$ on $(0,1)$ the sequence $\{\delta_m\}_{m=1}^{\infty}$ in Theorem 6.1 is unique, with*

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

6.3 General representation

Given a continuous distribution function $G(x)$ with support $\Xi \subset \mathbb{R}$, any distribution function $F(x)$ with support contained in Ξ can be written as

$F(x) = H(G(x))$, where $H(u) = F(G^{-1}(u))$ is a distribution function on $[0,1]$. Moreover, if F and G are absolutely continuous with densities f and g, respectively, then H is absolutely continuous with density $h(u)$, and $f(x) = h(G(x))g(x)$. Therefore, $f(x)$ can be estimated semiparametrically by estimating $h(u)$ semiparametrically.

In general, the role of the a priori chosen distribution function G is threefold:

1. G specifies the support of the unknown distribution functions F in the semi-nonparametric model;

2. G maps the parameter space $\mathcal{F}$ of candidate distributions for F one-to-one onto a space $\mathcal{H}(0,1)$ of distribution functions on the unit interval, which enables us to develop a unified inference approach for a wide range of semi-nonparametric models;

3. G serves as an initial guess for $F(x) = H(G(x))$. If the guess is right then $H(u) = u$. A related interpretation of G is that it serves as a (non-Bayesian) "prior" for F, with the estimate $\widehat{H}$ of H playing the role of correction mechanism which converts the prior G into a "posterior" $\widehat{F}$ for F on the basis of data evidence. Another related interpretation is that $F = G$ represents a standard parametric model for which the semi-nonparametric model is a generalization.

It follows now from Theorem 6.1 and (6.22) that

Theorem 6.3. *Given an absolutely continuous distribution function $G(x)$ on $\mathbb{R}$ with density $g(x)$, any density function $f(x)$ with support contained in the support of g* (i.e., $\{x : f(x) > 0\} \subset \{x : g(x) > 0\}$) *can be written as*

$$f(x) = g(x)\frac{\left(1 + \sqrt{2}\sum_{k=1}^{\infty} \delta_k \cos\left(k\pi G\left(x\right)\right)\right)^2}{1 + \sum_{m=1}^{\infty} \delta_m^2} \tag{6.14}$$

a.e. on $\{x : f(x) > 0\}$. Moreover, the corresponding SNP densities

$$f_n(x) = g(x)\frac{\left(1 + \sqrt{2}\sum_{k=1}^{n} \delta_k \cos\left(k\pi G\left(x\right)\right)\right)^2}{1 + \sum_{m=1}^{n} \delta_m^2} \tag{6.15}$$

satisfy

$$\int_{-\infty}^{\infty} |f(x) - f_n(x)|\, dx \leq \sqrt{5 \sum_{k=n+1}^{\infty} \delta_k^2} \to 0.$$

Furthermore, the SNP distribution function $F_n(x) = H_n(G(x))$ *satisfies*

$$\sup_x |F(x) - F_n(x)| \leq \sqrt{5 \sum_{k=n+1}^{\infty} \delta_k^2} \to 0,$$

where $F(x) = \int_{-\infty}^{x} f(z)dz$ *and* $H_n(u)$ *is defined by* (6.8).

6.4 Smoothness

The non-Euclidean parameter of a semi-nonparametric econometric model often takes the form of a density function $f(x)$. Usually it is assumed that $f(x)$ has certain smoothness and regularity properties, like boundedness, continuity and differentiability. Also, usually the semiparametric model involved requires that the support of $f(x)$ is connected, i.e.,

$$\{x \in \mathbb{R} : f(x) > 0\} = (a, b),$$

where possibly $a = -\infty$ and/or $b = \infty$. To impose these conditions, we need to impose corresponding smoothness and regularity conditions on the density $g(x)$ of the a priori chosen distribution function $G(x)$ and on the density $h(u)$ in the transformation $f(x) = h(G(x))g(x)$.

Denoting $u = G(x)$, we can write

$$h(u) = \frac{f(G^{-1}(u))}{g(G^{-1}(u))}$$

Given that G is chosen such that f and g have the same support (a, b), it follows that $h(u)$ must have support $(0, 1)$, i.e.,

$$h(u) > 0 \text{ on } (0, 1), \tag{6.16}$$

and if f and g are continuous on (a, b) then $h(u)$ is continuous on $(0, 1)$. Moreover, if it is known that $f(x) < \infty$ for each $x \in (a, b)$, then $f(x)/g(x) < \infty$ for each $x \in (a, b)$, hence $h(u) < \infty$ for each $u \in (0, 1)$. Furthermore, since $g(x)$ is an initial guess of $f(x)$, it is reasonable to assume that $g(x)$ is sufficiently close to $f(x)$ to guarantee that

$$\lim_{x\downarrow a} f(x)/g(x) < \infty,\ \lim_{x\uparrow b} f(x)/g(x) < \infty$$

These tail conditions, together with the condition that $f(x) < \infty$ for each $x \in (a,b)$, are equivalent to $\sup_{0\leq u\leq 1} h(u) < \infty$. A sufficient condition for the latter is that the δ_k's in (6.4) satisfy

$$\sum_{k=1}^{\infty} |\delta_k| < \infty. \tag{6.17}$$

This condition is stronger than necessary for $\sup_{0\leq u\leq 1} h(u) < \infty$ only, because:

Theorem 6.4. *Condition* (6.17) *implies that* $\sum_{k=1}^{\infty} \delta_k\sqrt{2}\cos(k\pi u)$ *is uniformly continuous on* $[0,1]$, *hence the corresponding density function* $h(u)$ *in* (6.4) *is then uniformly continuous on* $[0,1]$.[1]

Note that Theorems 6.2 and 6.4 imply the following corollary.

Theorem 6.5. *Suppose that* $h(u)$ *has support* $(0,1)$. *If the* δ_k*'s in* (6.4) *are confined to those for which* $\sum_{k=1}^{\infty} |\delta_k| < \infty$ *then they are unique.*

Next, suppose that f and g are continuously differentiable on (a,b). Then $h(u)$ is continuously differentiable on $(0,1)$. A sufficient condition for the latter is that

$$\sum_{k=1}^{\infty} k|\delta_k| < \infty. \tag{6.18}$$

To see this, pick any $u \in [0,1]$ and let $\varepsilon \neq 0$ be so small that $u+\varepsilon \in [0,1]$. Then by the mean value theorem there exists a sequence $\lambda_k(u,\varepsilon) \in [0,1]$ such that

$$\limsup_{\varepsilon\to 0}\left|\frac{1}{\varepsilon}\sum_{k=1}^{\infty}\delta_k\left(\cos(k\pi(u+\varepsilon)) - \cos(k\pi u)\right) + \pi\sum_{k=1}^{\infty} k\delta_k \sin(k\pi u)\right|$$

[1] Which implies that $\sup_{0\leq u\leq 1} h(u) < \infty$.

$$\leq \pi \limsup_{\varepsilon \to 0} \sum_{k=1}^{\infty} k|\delta_k|. |\sin(k\pi u) - \sin(k\pi(u + \lambda_k(u, \varepsilon)\varepsilon))|$$

$$\leq \pi \limsup_{\varepsilon \to 0} \sum_{k=1}^{n} k|\delta_k|. |\sin(k\pi u) - \sin(k\pi(u + \lambda_k(u, \varepsilon)\varepsilon))| + 2\pi \sum_{k=n+1}^{\infty} k|\delta_k|$$

$$= 2\pi \sum_{k=n+1}^{\infty} k|\delta_k| \to 0 \text{ as } n \to \infty,$$

hence

$$\frac{d}{du}\left(\sum_{k=1}^{\infty} \delta_k \cos(k\pi u)\right) = \sum_{k=1}^{\infty} \delta_k \frac{d\cos(k\pi u)}{du} = -\pi \sum_{k=1}^{\infty} k\delta_k \sin(k\pi u)$$

and thus

$$h'(u) = -\frac{2\pi\left(1 + \sqrt{2}\sum_{k=1}^{\infty} \delta_k \sqrt{2}\cos(k\pi u)\right)\left(\sum_{k=1}^{\infty} k\delta_k \sqrt{2}\sin(k\pi u)\right)}{1 + \sum_{i=1}^{\infty} \delta_i^2}.$$

Note that $h'(0) = h'(1) = 0$. Moreover, it follows similar to Theorem 6.4 that $h'(u)$ is uniformly continuous on $[0, 1]$.

Along the same lines it can be shown that

Theorem 6.6. *If for some natural number $\ell \geq 1$, $\sum_{k=1}^{\infty} k^{\ell}|\delta_k| < \infty$, then the density function $h(u)$ in (6.4) is ℓ-times continuously differentiable on $[0, 1]$.*

6.5 Bivariate densities

Similar to (4.30) and (6.1), any bivariate density $h(u, v)$ on $[0, 1] \times [0, 1]$ can be written as

$$h(u, v) = \left(\alpha_{0,0} + \sum_{k=1}^{\infty} \alpha_{k,0}\sqrt{2}\cos(k\pi u) + \sum_{m=1}^{\infty} \alpha_{0,m}\sqrt{2}\cos(m\pi v) + \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{k,m}\sqrt{2}\cos(k\pi u)\sqrt{2}\cos(m\pi v)\right)^2$$

a.e. on $[0, 1] \times [0, 1]$, where $\sum_{k=0}^{\infty}\sum_{m=0}^{\infty} \alpha_{k.m} = 1$, and similar to (6.4) we can reparametrize the $\alpha_{k,m}$'s such that $h(u, v)$ becomes

$$h(u, v) = \frac{1}{1 + \sum_{k=1}^{\infty} \delta_{k,0}^2 + \sum_{m=1}^{\infty} \delta_{0,m}^2 + \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \delta_{k,m}^2}$$

$$\times \left(1+\sum_{k=1}^{\infty}\delta_{k,0}\sqrt{2}\cos(k\pi u)+\sum_{m=1}^{\infty}\delta_{0,m}\sqrt{2}\cos(m\pi v)\right.$$
$$\left.+\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\delta_{k,m}\sqrt{2}\cos(k\pi u)\sqrt{2}\cos(m\pi v)\right)^2$$

a.e. on $[0,1]\times[0,1]$, where

$$\sum_{k=1}^{\infty}\delta_{k,0}^2+\sum_{m=1}^{\infty}\delta_{0,m}^2+\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\delta_{k,m}^2<\infty. \tag{6.19}$$

Note that the marginal densities of $h(u,v)$ take the form of a weighted sum of univariate densities. In particular, denoting

$$\begin{aligned} h_1(u) &= \int_0^1 h(u,v)dv \\ h_{1,0}(u) &= \frac{1+\sqrt{2}\sum_{k=1}^{\infty}\delta_{k,0}\cos(k\pi u)}{1+\sum_{k=1}^{\infty}\delta_{k,0}^2} \\ h_{1,m}(u) &= \frac{\left(\delta_{0,m}+\sum_{k=1}^{\infty}\delta_{k,m}\sqrt{2}\cos(k\pi u)\right)^2}{\sum_{k=0}^{\infty}\delta_{k,m}^2} \end{aligned}$$

it can be shown that

$$h_1(u)=\frac{\left(1+\sum_{k=1}^{\infty}\delta_{k,0}^2\right)h_{1,0}(u)+\sum_{m=1}^{\infty}\left(\sum_{k=0}^{\infty}\delta_{k,m}^2\right)h_{1,m}(u)}{1+\sum_{k=1}^{\infty}\delta_{k,0}^2+\sum_{m=1}^{\infty}\delta_{0,m}^2+\sum_{k=1}^{\infty}\sum_{m=1}^{\infty}\delta_{k,m}^2}.$$

Of course, the $\delta_{k,m}$'s can be reparametrized such that $h_1(u)$ takes the form (6.4).

Similar to (6.6), let

$$\begin{aligned} h_n(u,v) &= \frac{1}{1+\sum_{k=1}^{n}\delta_{k,0}^2+\sum_{m=1}^{n}\delta_{0,m}^2+\sum_{k=1}^{n}\sum_{m=1}^{n}\delta_{k,m}^2} \\ &\quad\times\left(1+\sum_{k=1}^{n}\delta_{k,0}\sqrt{2}\cos(k\pi u)+\sum_{m=1}^{n}\delta_{0,m}\sqrt{2}\cos(m\pi v)\right. \\ &\quad\left.+\sum_{k=1}^{n}\sum_{m=1}^{n}\delta_{k,m}\sqrt{2}\cos(k\pi u)\sqrt{2}\cos(m\pi v)\right)^2 \end{aligned}$$

be the truncated version of $h(u,v)$. It is not hard to verify that

$$\lim_{n\to\infty}\int_0^1\int_0^1 |h(u,v)-h_n(u,v)|\,dudv=0.$$

Finally, note that any density $f(x,y)$ with support $\Xi_x\times\Xi_y\subset\mathbb{R}^2$ can be represented by

$$f(x,y)=g_x(x)g_y(y)h\left(G_x(x),G_y(y)\right)\text{ a.e. on }\Xi_x\times\Xi_y,$$

with truncated version

$$f_n(x,y)=g_x(x)g_y(y)h_n\left(G_x(x),G_y(y)\right),$$

where G_x is an a priori chosen absolutely continuous distribution function with density g_x and support Ξ_x, and G_y is an a priori chosen absolutely continuous distribution function with density g_y and support Ξ_y.

6.6 Appendix: Proofs

6.6.1 Theorem 6.1

The result (6.8) follows from the well-known sine-cosine formulas. To prove (6.7), denote

$$\begin{aligned} f(u) &= \frac{1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos(k\pi u)}{\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2}}, \\ f_n(u) &= \frac{1+\sum_{k=1}^{n}\delta_k\sqrt{2}\cos(k\pi u)}{\sqrt{1+\sum_{m=1}^{n}\delta_m^2}} \end{aligned}$$

It follows from the Cauchy-Schwarz inequality that

$$\begin{aligned} \int_0^1 \left|f(u)^2-f_n(u)^2\right|du &= \int_0^1 |f(u)-f_n(u)|\,.\,|f(u)+f_n(u)|\,du \\ &\leq 2\sqrt{\int_0^1 (f(u)-f_n(u))^2\,du} \end{aligned} \tag{6.20}$$

Moreover,

$$
\begin{aligned}
&\int_0^1 \left(f\left(u\right)-f_n\left(u\right)\right)^2 du \\
&= \left(1+\sum_{k=1}^{n}\delta_k^2\right)\left(\frac{1}{\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2}}-\frac{1}{\sqrt{1+\sum_{m=1}^{n}\delta_m^2}}\right)^2 \\
&+\frac{\sum_{k=n+1}^{\infty}\delta_k^2}{1+\sum_{m=1}^{\infty}\delta_m^2} \leq \frac{5}{4}\sum_{k=n+1}^{\infty}\delta_k^2 \qquad (6.21)
\end{aligned}
$$

as is not hard to verify. The result (6.7) now follows from (6.20) and (6.21). Finally, (6.9) follows from

$$
\sup_{0\leq u\leq 1}\left|H(u)-H_n\left(u\right)\right| \leq \int_0^1 \left|h\left(x\right)-h_n\left(x\right)\right| dx \leq \sqrt{5\sum_{k=n+1}^{\infty}\delta_k^2} \to 0. \qquad (6.22)
$$

6.6.2 Theorem 6.4

Pick any $u\in[0,1]$ and let $\varepsilon\neq 0$ be so small that $u+\varepsilon\in[0,1]$. Then

$$
\begin{aligned}
&\limsup_{\varepsilon\to 0}\left|\sum_{k=1}^{\infty}\delta_k\left(\cos(k\pi(u+\varepsilon))-\cos(k\pi u)\right)\right| \\
&=\limsup_{\varepsilon\to 0}\left|\sum_{k=1}^{\infty}\delta_k\left((\cos(k\pi\varepsilon)-1)\cos(k\pi u)-\sin(k\pi\varepsilon)\sin(k\pi u)\right)\right| \\
&\leq \limsup_{\varepsilon\to 0}\sum_{k=1}^{n}|\delta_k|\left(|1-\cos(k\pi\varepsilon)|+|\sin(k\pi\varepsilon)|\right)+3\sum_{k=n+1}^{\infty}|\delta_k| \\
&=3\sum_{k=n+1}^{\infty}|\delta_k| \to 0 \text{ as } n\to\infty.
\end{aligned}
$$

By the compactness of $[0,1]$ this result implies that $\sum_{k=1}^{\infty}\delta_k\cos(k\pi u)$ is uniformly continuous on $[0,1]$, and so is $h(u)$.

Chapter 7

Compactness

As said before, the non-Euclidean parameter of a semi-nonparametric econometric model often takes the form of a density and/or distribution function. See the next section for an example. Similar to parametric nonlinear estimation, these non-Euclidean parameters need to be confined to a compact metric space. In this chapter it will be shown how to construct such compact metric spaces.

7.1 General density and distribution functions

Recall that the results in Theorem 6.1 read more generally as follows. Given a complete orthonormal sequence $\{\rho_k\}_{k=0}^{\infty}$ in $L^2(0,1)$ with $\rho_0(u) \equiv 1$, for every density function $h(u)$ on $[0,1]$ there exist uncountable many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying

$$\sum_{m=1}^{\infty} \delta_m^2 < \infty \tag{7.1}$$

such that

$$h(u) = \frac{\left(1 + \sum_{m=1}^{\infty} \delta_m \rho_m(u)\right)^2}{1 + \sum_{m=1}^{\infty} \delta_m^2} \text{ a.e.} \tag{7.2}$$

Moreover, recall that this representation does not require any smoothness conditions. Thus (7.2) holds if $h(u)$ is merely Borel measurable. Furthermore, denoting

$$h_n(u) = \frac{\left(1 + \sum_{m=1}^{n} \delta_m \rho_m(u)\right)^2}{1 + \sum_{m=1}^{n} \delta_m^2} \tag{7.3}$$

for $n \geq 1$, it follows that

$$\int_0^1 |h(u) - h_n(u)|\, du \leq \sqrt{5 \sum_{k=n+1}^{\infty} \delta_k^2} \to 0 \tag{7.4}$$

as $n \to \infty$.

The condition (7.1) can be imposed by imposing the restrictions $|\delta_k| \leq \overline{\delta}_k$, where $\overline{\delta}_k$ is an a priori chosen positive sequence such that $\sum_{k=1}^{\infty} \overline{\delta}_k^2 < \infty$. For example, let

$$\overline{\delta}_k = \frac{c}{1 + \sqrt{k}\ln(k)}, \tag{7.5}$$

for some constant $c > 0$. It is easy to verify that then $\sum_{k=1}^{\infty} \overline{\delta}_k^2 < c^2 + c^2/\ln(2)$.

These restrictions on the δ_k's also play a key-role in proving compactness:

Theorem 7.1. *Let $\mathcal{D}(0,1)$ be the space of densities of the type* (7.2) *subject to the restrictions $|\delta_k| \leq \overline{\delta}_k$ for some a priori chosen positive sequence $\overline{\delta}_k$ satisfying $\sum_{k=1}^{\infty} \overline{\delta}_k^2 < \infty$, endowed with the L^1 metric*

$$\|h_1 - h_2\|_1 = \int_0^1 |h_1(u) - h_2(u)|\, du.$$

Then $\mathcal{D}(0,1)$ is compact. Consequently, the space

$$\mathcal{H}(0,1) = \left\{ H(u) = \int_0^u h(v)dv,\ h \in \mathcal{D}(0,1) \right\}$$

endowed with the "sup" metric

$$\|H_1 - H_2\|_{\sup} = \sup_{0 \leq u \leq 1} |H_1(u) - H_2(u)|$$

is compact as well. Moreover, let $\mathcal{D}_n(0,1)$ be the space of SNP densities of the type (7.3)*, with $\mathcal{D}_0(0,1)$ the singleton $\{h(u) \equiv 1\}$, subject to the same restrictions on the δ_k's, and endowed with the same metric as $\mathcal{D}(0,1)$. Then the sequence $\mathcal{D}_n(0,1)$ is dense in $\mathcal{D}(0,1)$:*

$$\mathcal{D}(0,1) = \overline{\cup_{n=0}^{\infty} \mathcal{D}_n(0,1)}.$$

Consequently, the spaces

$$\mathcal{H}_n(0,1) = \left\{ H_n(u) = \int_0^u h_n(v)dv,\ h_n \in \mathcal{D}_n(0,1) \right\}$$

endowed with the sup metric are dense in $\mathcal{H}(0,1)$:

$$\mathcal{H}(0,1) = \overline{\cup_{n=0}^{\infty}\mathcal{H}_n(0,1)}.$$

Bierens (2008, Theorem 8) proved this result for the case where the $\rho_m(u)$ are Legendre polynomials and $\overline{\delta}_k$ is given by (7.5). However, as will be shown below these results hold for any complete orthonormal sequence $\rho_n(u)$ and any positive sequence $\overline{\delta}_k$ satisfying $\sum_{k=1}^{\infty}\overline{\delta}_k^2 < \infty$.

Similarly, it is easy to construct compact metric spaces of general density and distribution functions on $\mathbb{R}$. In particular, recall that any density $f(x)$ with support $\mathbb{X} \subset \mathbb{R}$ can be written as $f(x) = h(G(x))g(x)$, where $G(x)$ is a given absolutely continuous distribution function with density $g(x)$ and support containing $\mathbb{X}$: $\mathbb{X} \subset \{x \in \mathbb{R} : g(x) > 0\}$. Thus, denoting

$$\begin{aligned}\mathcal{D}(G) &= \{f(x) = h(G(x))g(x) : h \in \mathcal{D}(0,1)\} \\ \mathcal{F}(G) &= \left\{F(x) = \int_{-\infty}^{x} f(z)dz : f \in \mathcal{D}(G)\right\}\end{aligned}$$

it follows trivially from Theorem 7.1 that $\mathcal{D}(G)$ is a compact metric space of densities with metric $\int_{-\infty}^{\infty}|f_1(x) - f_2(x)|\,dx$, and $\mathcal{F}(G)$ is a compact metric space of absolutely continuous distribution functions with metric $\sup_{x\in\mathbb{R}}|F_1(x) - F_2(x)|$. Moreover, denoting

$$\begin{aligned}\mathcal{D}_n(G) &= \{f(x) = h(G(x))g(x) : h \in \mathcal{D}_n(0,1)\} \\ \mathcal{F}_n(G) &= \left\{F(x) = \int_{-\infty}^{x} f(z)dz : f \in \mathcal{D}_n(G)\right\}\end{aligned}$$

it follows from Theorem 7.1 that $\mathcal{D}(G) = \overline{\cup_{n=0}^{\infty}\mathcal{D}_n(G)}$ and $\mathcal{F}(G) = \overline{\cup_{n=0}^{\infty}\mathcal{F}_n(G)}$.

The compactness part of Theorem 7.1 follows from the following two lemmas and the fact that similar to (6.7), for each pair $h_1, h_2 \in \mathcal{D}(0,1)$ there exist sequences $\delta_1 = \{\delta_{1,k}\}_{k=1}^{\infty}$ and $\delta_2 = \{\delta_{2,k}\}_{k=1}^{\infty}$ such that

$$\int_0^1 |h_1(u) - h_2(u)|\,du \le \sqrt{5}\sqrt{\sum_{k=1}^{\infty}(\delta_{1,k} - \delta_{2,k})^2}.$$

Lemma 7.1. *Let $\{\overline{\delta}_k\}_{k=1}^{\infty}$ be an a priori chosen positive sequence satisfying $\sum_{k=1}^{\infty}\overline{\delta}_k^2 < \infty$, and let $\Delta = \times_{k=1}^{\infty}[-\overline{\delta}_k, \overline{\delta}_k]$. Endow the space Δ with the metric*

$$d(\delta_1, \delta_2) = \sqrt{\sum_{k=1}^{\infty}(\delta_{1,k} - \delta_{2,k})^2},$$

where $\delta_1 = \{\delta_{1,k}\}_{k=1}^{\infty} \in \Delta,\ \delta_2 = \{\delta_{2,k}\}_{k=1}^{\infty} \in \Delta$. Then Δ is compact.

Lemma 7.2. *Let $s(\delta_1, \delta_2)$ be another metric on Δ such that for some constant $c > 0$, $s(\delta_1, \delta_2) \leq c.d(\delta_1, \delta_2)$. Then under the conditions of Lemma 7.2, the space Δ endowed with the metric s is compact as well.*

7.2 Smooth densities on the unit interval

Note that if we replace the condition $\sum_{k=1}^{\infty}\overline{\delta}_k^2 < \infty$ in Lemma 7.1 by $\sum_{k=1}^{\infty} k^{\ell}\overline{\delta}_k < \infty$ for some integer $\ell \geq 0$ and the metric $d(\delta_1, \delta_2)$ by

$$d(\delta_1, \delta_2) = \sum_{k=1}^{\infty} k^{\ell}|\delta_{1,k} - \delta_{2,k}|$$

then the result of Lemma 7.1 carries over. Consequently, the following results hold.

Theorem 7.2. *Let $\mathcal{D}_{\ell}(0,1)$ be the space of densities of the type* (6.4) *subject to the restrictions $|\delta_k| \leq \overline{\delta}_k$ for some a priori chosen positive sequence $\overline{\delta}_k$ satisfying $\sum_{k=1}^{\infty} k^{\ell}\overline{\delta}_k < \infty$ for some integer $\ell \geq 0$. Endow $\mathcal{D}_{\ell}(0,1)$ with the Sobolev*[1] *metric*

$$\|h_1 - h_2\|_{\ell} = \max_{0 \leq m \leq \ell} \sup_{0 \leq u \leq 1} \left| h_1^{(m)}(u) - h_2^{(m)}(u) \right|, \tag{7.6}$$

where $h^{(m)}(u) = d^m h(u)/(du)^m$ for $m \geq 1$, $h^{(0)}(u) = h(u)$. Then $\mathcal{D}_{\ell}(0,1)$ is compact. Moreover, let $\mathcal{D}_{\ell,n}(0,1)$ be the space of SNP densities of the type (6.6), *subject to the same restrictions on the δ_k's, and endowed with the same metric as $\mathcal{D}_{\ell}(0,1)$. Again, $\mathcal{D}_{\ell,0}(0,1)$ is the singleton $\{h(u) \equiv 1\}$. Then the sequence $\mathcal{D}_{\ell,n}(0,1)$ is dense in $\mathcal{D}_{\ell}(0,1)$: $\mathcal{D}_{\ell}(0,1) = \overline{\cup_{n=0}^{\infty}\mathcal{D}_{\ell,n}(0,1)}$.*

[1]See for example Adams and Fournier (2003).

This result follows from the fact that for each pair $h_1, h_2 \in \mathcal{D}_\ell(0,1)$ with corresponding sequences $\{\delta_{1,k}\}_{k=1}^{\infty}$ and $\{\delta_{2,k}\}_{k=1}^{\infty}$ we have

$$\|h_1 - h_2\|_\ell = O\left(\sum_{k=1}^{\infty} k^\ell |\delta_{1,k} - \delta_{2,k}|\right), \tag{7.7}$$

as is not hard to verify.

7.3 Appendix: Proofs

7.3.1 Lemma 7.1

To prove the compactness of Δ it suffices to prove that Δ is complete and totally bounded. See Royden (1968, Proposition 15, p.164).

Completeness means that every Cauchy sequence in Δ takes a limit in Δ. To show this, let $\delta_n = \{\delta_{n,k}\}_{k=1}^{\infty}$ be an arbitrary Cauchy sequence in Δ, i.e.,

$$\lim_{\min(n,m)\to\infty} d(\delta_n, \delta_m) = \lim_{\min(n,m)\to\infty} \sqrt{\sum_{k=1}^{\infty} (\delta_{n,k} - \delta_{m,k})^2} = 0.$$

Then for each $k \geq 1$, $\lim_{\min(n,m)\to\infty} |\delta_{n,k} - \delta_{m,k}| = 0$, hence $\delta_{n,k}$ is a Cauchy sequence in $[-\overline{\delta}_k, \overline{\delta}_k]$ and therefore takes a limit $\delta_k \in [-\overline{\delta}_k, \overline{\delta}_k]$. Consequently, $\delta = \{\delta_k\}_{k=1}^{\infty} \in \Delta$ and $\lim_{n\to\infty} d(\delta_n, \delta) = 0$, where the latter follows from

$$\begin{aligned}
\limsup_{n\to\infty} (d(\delta_n, \delta))^2 &= \limsup_{n\to\infty} \sum_{k=1}^{\infty} (\delta_{n,k} - \delta_k)^2 \\
&\leq \limsup_{n\to\infty} \sum_{k=1}^{m} (\delta_{n,k} - \delta_k)^2 + 4 \sum_{k=m+1}^{\infty} \overline{\delta}_k^2 \\
&= 4 \sum_{k=m+1}^{\infty} \overline{\delta}_k^2 \to 0 \text{ as } m \to \infty.
\end{aligned}$$

Thus, Δ is complete.

To prove total boundedness, let $\varepsilon > 0$ be arbitrary, and choose an n so large that $\sqrt{\sum_{k=n+1}^{\infty} \overline{\delta}_k^2} < \varepsilon/4$. Denote

$$\Delta_n = \left(\mathsf{X}_{k=1}^{n}[-\overline{\delta}_k, \overline{\delta}_k]\right) \times \left(\mathsf{X}_{k=n+1}^{\infty}\{0\}\right). \tag{7.8}$$

Since $\times_{k=1}^{n}[-\overline{\delta}_k, \overline{\delta}_k]$ is a closed and bounded subset of $\mathbb{R}^n$ it is compact, hence Δ_n is compact. Therefore, there exist elements $\delta_1, ..., \delta_M$ of Δ_n such that $\Delta_n \subset \cup_{i=1}^{M}\{\delta_* \in \Delta_n : d(\delta, \delta_i) < \varepsilon/2\}$. Since for each $\delta \in \Delta$ there exists a $\delta_* \in \Delta_n$ such that $d(\delta, \delta_*) \leq 2\sqrt{\sum_{k=n+1}^{\infty} \overline{\delta}_k^2} < \varepsilon/2$, it follows now that each $\delta \in \Delta$ belongs to one of the open sets $\{\delta \in \Delta : d(\delta, \delta_i) < \varepsilon\}$, hence $\Delta \subset \cup_{i=1}^{M}\{\delta \in \Delta : d(\delta, \delta_i) < \varepsilon\}$. Thus, Δ is totally bounded.

7.3.2 Lemma 7.2

Let Δ_{O} be a set which is open under the metric $s(.,.)$ but not under the metric $d(.,.)$. Let $\underline{\delta} \in \Delta_{\mathrm{O}}$ be a point of closure under the d-metric. Note that by assumption, $\underline{\delta}$ is an interior point of Δ_{O} under the s-metric. Then for every $\varepsilon > 0$ there exists a $\delta \notin \Delta_{\mathrm{O}}$ such that $d(\underline{\delta}, \delta) < \varepsilon$. But then $s(\underline{\delta}, \delta) < \varepsilon/c$, which would imply that $\underline{\delta}$ is a point of closure under the s-metric as well. This contradiction implies that open sets under the s-metric are also open under the d-metric. Consequently, any open covering of Δ under the s-metric is an open covering under the d-metricy. Since in the latter case Δ is compact, there exists a finite sub-covering of Δ, which is also a finite sub-covering under the s-metric. Hence Δ is compact under the s-metric.

Part II

Semi-Nonparametric models (To be done)

References

Adams, R.A. & J.J.F. Fournier (2003), *Sobolev Spaces.* Academic Press.

Andrews, D.W.K. (1994), "Asymptotics for Semiparametric Econometric Models Via Stochastic Equicontinuity", *Econometrica* 62, 43-72.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov & J.A. Wellner (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, Springer.

Bierens, H. J. (1997), "Testing the Unit Root with Drift Hypothesis Against Nonlinear Trend Stationarity, with an Application to the U.S. Price Level and Interest Rate", *Journal of Econometrics* 81, 29-64.

Bierens, H.J. (2004), *Introduction to the Mathematical and Statistical Foundations of Econometrics.* Cambridge University Press.

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

Bierens, H.J. (2011), "Consistency and Asymptotic Normality of Sieve Estimators Under Weak and Verifiable Conditions". http://econ.la.psu.edu/~hbierens/SNPMODELS.PDF

Bierens, H.J. & J. R. Carvalho (2007), "Semi-Nonparametric Competing Risks Analysis of Recidivism", *Journal of Applied Econometrics* 22, 971-993.

Bierens, H. J. & L. F. Martins (2010), "Time Varying Cointegration", *Econometric Theory* 26, 1453-1490.

Bierens, H.J. & and H. Pott-Buter (1990), "Specification of Engel Curves by Nonparametric Regression (with discussion)", *Econometric Reviews* 9, 123-184.

Billingsley, P. (1968), *Convergence of Probability Measures.* John Wiley.

Chen, X. (2007), "Large sample sieve estimation of semi-nonparametric models". In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics, Vol. 6,* Ch. 76. Elsevier.

Chen, X., O. Linton & I. Van Keilegom (2003), "Estimation of Semiparametric Models when the Criterion Function is Not Smooth", *Econometrica*, 71, 1591-1608.

Eastwood, B.J. & A.R. Gallant (1991), "Adaptive Rules for Semi-Nonparametric Estimators that Achieve Asymptotic Normality", *Econometric Theory* 7, 307-340.

Elbers, C. & G. Ridder (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model", *Review of Economic Studies* 49, 403-409.

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

Gabler, S, F. Laisney & M. Lechner (1993), "Seminonparametric Estimation of Binary-Choice Models with an Application to Labor-Force Participation", *Journal of Business & Economic Statistics* 11, 61-80.

Gallant, A.R. & D.W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica* 55, 363-390.

Gill, R.D. (1989), "Non- and Semi-Parametric Maximum Likelihood Estimators and the Von Mises Method (Part 1)", *Scandinavian Journal of Statistics* 16, 97-128.

Grenander, U. (1981), *Abstract Inference.* Wiley.

Hamming, R.W. (1973), *Numerical Methods for Scientists and Engineers.* Dover Publications.

Hannan, E.J., & B.G. Quinn (1979), "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society, Series B,* 41, 190-195.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161.

Horowitz, J.L. (1998), *Semiparametric Methods in Econometrics*, Springer.

Jennrich, R.I. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-643.

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

Lancaster, T. (1979), "Econometric Methods for the Duration of Unemployment", *Econometrica* 47, 939-956.

Manski, C.F. (1988), "Identification of Binary Response Models", *Journal of the American Statistical Association* 83, 729-738.

Newey, W.K. (1997), "Convergence Rates and Asymptotic Normality for Series Estimators", *Journal of Econometrics* 79, 147-168.

Royden, H.L. (1968), *Real Analysis.* Macmillan.

Schwarz, G. (1978), "Estimating the Dimension of a Model", *Annals of Statistics* 6, 461-464.

Young, N. (1988), *An Introduction to Hilbert Space.* Cambridge University Press.

Shen, X. (1997), "On the Method of Sieves and Penalization", *Annals of Statistics* 25, 2555-2591.

Sims, C.A. (1980), ”Macroeconomics and Reality”, *Econometrica* 48, 1-48.

White, H. & J. Wooldridge (1991), ”Some Results on Sieve Estimation with Dependent Observations”. In W.A. Barnett, J. Powell & G. Tauchen (eds.), *Non-parametric and Semi-parametric Methods in Econometrics and Statistics*, Ch. 18, Cambridge University Press.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series.* Almqvist and Wiksell, Sweden.