# Introduction to Hilbert Space Theory and Its Econometric Applications

Herman J. Bierens

Professor Emeritus of Economics

Pennsylvania State University

Current version: June 29, 2018

# 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).

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

Gallant and Nychka (1987) consider SNP estimation of Heckman's (1979) sample selection model, where the bivariate error distribution of the latent 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

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

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. In chapter 4 I will discuss the famous Wold (1938) decomposition theorem, which will be derived first in general terms and then for covariance stationary time series. Also, in this chapter 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), I will review in chapter 5 the various ways one can construct orthonormal polynomials that span a given Hilbert space of functions. In chapter 6 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 7, 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 8 I will derive condition under which the SNP density based on the cosine series, and their first, second and third derivatives converges uniformly on $[0,1]$ to the true density and its derivatives, respectively. In chapter 9 I will introduce the SNP Tobit model, and establish mild conditions for its SNP identification. Moreover, in this chapter I will establish low-level conditions for the consistency of the sieve ML estimators of the parameters involved.

In chapter 10 I will discuss Shen's (1997) approach for deriving asymptotic normality results for sieve ML estimators, which is nowadays the standard in the sieve ML estimation literature.

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 $I(.)$. Furthermore, the bold-face $\mathbf{i}$ denotes the complex number $\mathbf{i} = \sqrt{-1}$.

Finally, chapters 6 and 8 come each with 7 pictures. In this manuscript only the picture captions involved are listed. The actual images can be viewed from:
http://www.personal.psu.edu/hxb11/HILBERTSPACE_PICTURES.PDF

## Acknowledgment

The helpful comments and suggestions of Flor Gabrielli (Conicet-UNCuyo, Argentina) are gratefully acknowledged. Of course, all errors are mine.

# Part I

# Hilbert space theory

# 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 well 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]
$(e)$ $1.x = x$;
$(f)$ $(c_1 c_2).x = c_1.(c_2.x)$;
$(g)$ $c.(x + y) = c.x + c.y$;
$(h)$ $(c_1 + c_2).x = c_1.x + c_2.x$.

[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*[2] *$c$, $c.x$ 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 $x = (x_1..., x_n)'$ and $y = (y_1, ..., y_n)'$ is defined as $x'y = \sum_{i=1}^n x_i y_i$, which is a mapping $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ 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.

[2]Recal that a scalar is either a real number or a complex number.

**Definition 2.3**. *An inner product on a real vector space* $\mathcal{V}$ *is a real function* $\langle x, y\rangle$: $\mathcal{V} \times \mathcal{V} \to \mathbb{R}$ *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,* $\langle x, y\rangle$: $\mathcal{V} \times \mathcal{V} \to \mathbb{C}$. *Condition* (*1*) *then becomes*
(*1**) $\langle x, y\rangle = \overline{\langle y, x\rangle}$,[3]
*and* (*2*) *now holds for all complex and real numbers c. Note that also in this case* $\langle x, x\rangle$ *is real valued.*[4] *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 uniformly continuous real functions on $[0, 1]$, the integral $\langle f, g\rangle = \int_0^1 f(u)\, g(u)\mathrm{d}u$ is an inner product, with norm $\|f\| = \sqrt{\int_0^1 f(u)^2\, \mathrm{d}u}$. 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. Note that (2.2) also holds for complex scalars $c$.

[3]The bar denotes the complex conjugate: for $z = a + i.b$, $\overline{z} = a - i.b$.
[4]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

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

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 $\mathcal{V}$,

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

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

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

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

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 $\mathbb{R}$, endowed with the Hellinger distance

$$d(f,g) = \sqrt{\frac{1}{2}\int_{-\infty}^{\infty}\left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 \mathrm{d}x}.$$

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 of Euclidean spaces is still missing, namely that every Cauchy sequence in $\mathcal{V}$ 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 $n_0(\varepsilon) \in \mathbb{N}$ 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

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

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 uniformly 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 $\langle f,g\rangle = \int_0^1 f(u)g(u)\mathrm{d}u$ 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

$$||f_k - f_m||^2 = \int_0^1 (f_k(u) - f_m(u))^2 \mathrm{d}u < \frac{1}{3}\left(2^{-k} + 2^{-m}\right),$$

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

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 $\langle f,g\rangle = \int_0^1 f(u)g(u)\mathrm{d}u$ but also by $\langle f,g\rangle = \int_0^1 uf(u)g(u)\mathrm{d}u$, 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. This requirement 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 $\mathcal{H}$.*

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 $\min(k, m) \to \infty$.

Moreover, since a Hilbert space is a normed space (c.f. Definition 2.5), every point $z$ in a Hilbert space has a finite norm: $||z|| < \infty$. Therefore, implicit in the definition of a Hilbert space is the requirement that the inner product is chosen such that the implied norm is finite.

### 2.5.2 Banach spaces

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 therefore 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. See Definition 2.4 for the latter.

### 2.5.3 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 the 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 same inner product and associated norm and metric as $\mathcal{H}$.*

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 sequences 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.

### 2.5.4 Orthogonal complements

Given a sub-Hilbert space $\mathcal{S}$ of a Hilbert space $\mathcal{H}$, consider the space

$$\mathcal{S}^{\perp} = \{y \in \mathcal{H} : \langle x, y\rangle = 0 \text{ for all } x \in \mathcal{S}\}.$$

Such a space is called the *orthogonal complement* of $\mathcal{S}$, and is a Hilbert space itself:

**Lemma 2.1**. *Orthogonal complements are Hilbert spaces.*

### 2.5.5 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 2.2**. *$\mathcal{M}_m$ 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 2.3**. *$\mathcal{M}_\infty$ is a Hilbert space.*

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

The following related notions of completeness and separability play a key-role in semi-nonparametric modeling and inference:

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

**Remark**. There exist Hilbert spaces that are nonseparable.[5] However, all the Hilbert spaces considered in this monograph are separable, as separability is crucial for econometric applications.

Using the results in the next chapter it can be shown that $\mathcal{H}$ is separable if and only if there exists a complete orthonormal sequence in $\mathcal{H}$. Thus, the sequence $\{x_k\}_{k=1}^{\infty}$ in Definition 2.11 can then be chosen such that for all $k, m \in \mathbb{N}$, $\langle x_k, x_m \rangle = I(k = m)$.

## 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

The following two function spaces play a key-role in SNP modeling.

[5] Just Google "nonseparable Hilbert spaces" for examples.

**Definition 2.12**. *Given a probability density $w(x)$ on $\mathbb{R}$, let $L^2(w)$ be the space of Borel measurable real functions $f$ on $\mathbb{R}$ satisfying $\int_{-\infty}^{\infty} f(x)^2 w(x)\mathrm{d}x < \infty$, endowed with the inner product $\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g(x)w(x)\mathrm{d}x$ and associated norm $||f|| = \sqrt{\langle f, f\rangle}$ and metric $||f-g||$.*

**Definition 2.13**. *For $-\infty \leq a < b \leq \infty$, let $L^2(a,b)$ be the space of Borel measurable real functions $f$ on $\mathbb{R}$ satisfying $\int_a^b f(x)^2\mathrm{d}x < \infty$, endowed with the inner product $\langle f, g\rangle = \int_a^b f(x)g(x))\mathrm{d}x$ and associated norm $||f|| = \sqrt{\langle f, f\rangle}$ and metric $||f-g||$.*

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

**Theorem 2.5.** *$L^2(w)$ is a Hilbert space,*

and similarly,

**Theorem 2.6.** *$L^2(a,b)$ is a Hilbert space.*

As to the latter, let $w(x)$ be a density with support $(a,b)$, i.e., $w(x) > 0$ for all $x \in (a,b)$, $w(x) = 0$ for all $x \notin (a,b)$. Then any Cauchy sequence $f_n(x) \in L^2(a,b)$ corresponds to a Cauchy sequence $g_n(x) = f_n(x)/\sqrt{w(x)} \in L^2(w)$. It follows from Theorem 2.5 that there exists a $g \in L^2(w)$ such that

$$0 = \lim_{n\to\infty} \int_{-\infty}^{\infty} (g_n(x) - g(x))^2 w(x)\mathrm{d}x = \lim_{n\to\infty} \int_a^b (f_n(x) - f(x))^2 \mathrm{d}x,$$

where $f(x) = \sqrt{w(x)}g(x) \in L^2(a,b)$.

**Remark**. The condition in Definition 2.12 that $w(x)$ is a density on $\mathbb{R}$ may be replaced by the condition that $w(x)$ is a density on $\mathbb{R}^d$, simply by replacing the integral $\int_{-\infty}^{\infty}$ with $\int_{\mathbb{R}^d}$. Then Theorem 2.5 carries over, whereas Theorem 2.6 now reads: $L^2(\mathbb{R}^d)$ is a Hilbert space.

### 2.6.3 A Hilbert space of countable infinite sequences

Consider the following space of one-sided infinite sequences:

$$\mathbb{R}^{\infty} = \left\{ \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \delta_m^2 < \infty \right\}$$

For $\boldsymbol{\delta}_i = \{\delta_{i,m}\}_{m=1}^{\infty} \in \mathbb{R}^{\infty}$, $i = 1, 2$, endow this space with inner product

$$\langle \boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \rangle = \sum_{m=1}^{\infty} \delta_{1,m}\delta_{2,m}$$

and associated (pseudo-Euclidean) norm $||\boldsymbol{\delta}|| = \sqrt{\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle} = \sqrt{\sum_{m=1}^{\infty} \delta_m^2}$ and (pseudo-Euclidean) metric $||\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2||$. Then

**Theorem 2.7.** $\mathbb{R}^{\infty}$ *is a Hilbert space.*

## 2.7 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 $\lambda$,

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

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

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

Hence

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

The latter is minimal for

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

Substituting these solutions in (2.9) yields

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

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 $\varepsilon$ 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

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 $y \in \mathcal{S}^{\perp}$.

### 2.7.5 Lemma 2.2

Without loss of generality we may assume that the $m \times m$ matrix $\Sigma_m$ 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 $r = \text{rank}(\Sigma_m)$. 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 $c_j \in \mathbb{R}$. 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 $\mathcal{M}_m$.

### 2.7.6 Theorem 2.4

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

$$0 = \lim_{n\to\infty} \sup_{m\in\mathbb{N}} ||X_n - X_{n+m}||^2 = \lim_{n\to\infty} \sup_{m\in\mathbb{N}} E[(X_n - X_{n+m})^2]$$

hence by Chebyshev's inequality

$$\Pr[|X_n - X_{n+m}| > \varepsilon] \leq \frac{\sup_{\ell\in\mathbb{N}} E[(X_n - X_{n+\ell})^2]}{\varepsilon^2}$$

uniformly in $m \in \mathbb{N}$.

Taking $\varepsilon = k^{-2}$, $k \in \mathbb{N}$, and choosing $n = n_k$ such that $\sup_{\ell\in\mathbb{N}} E[(X_{n_k} - X_{n_k+\ell})^2] < k^{-6}$, it follows that $\sup_{m\in\mathbb{N}} \Pr[|X_{n_k} - X_{n_k+m}| > k^{-2}] \leq k^{-2}$, and the same applies if we replace $n_k + m$ by $n_{k+m}$. Hence for all $m \in \mathbb{N}$,

$$\sum_{k=1}^{\infty} \Pr[|X_{n_k} - X_{n_{k+m}}| > k^{-2}] \leq \sum_{k=1}^{\infty} k^{-2} < \infty$$

and therefore for arbitrary $\varepsilon > 0$,

$$0 = \lim_{\ell\to\infty} \sum_{k=\ell}^{\infty} \Pr[|X_{n_k} - X_{n_{k+m}}| > \varepsilon] \geq \lim_{\ell\to\infty} P(\widetilde{A}_{\ell,m}(\varepsilon))$$

where

$$\widetilde{A}_{\ell,m}(\varepsilon) = \cup_{k=\ell}^{\infty}\{\omega \in \Omega : |X_{n_k}(\omega) - X_{n_{k+m}}(\omega)| > \varepsilon\},$$

because $P(\widetilde{A}_{\ell,m}(\varepsilon)) \leq \sum_{k=\ell}^{\infty} \Pr[|X_{n_k} - X_{n_{k+m}}| > \varepsilon]$.

Denoting

$$A_{\ell,m}(\varepsilon) = \cap_{k=\ell}^{\infty}\{\omega \in \Omega : |X_{n_k}(\omega) - X_{n_{k+m}}(\omega)| \leq \varepsilon\}$$

we now have by the increasing monotonicity of $A_{\ell,m}(\varepsilon)$ in $\ell$ that for all $m \in \mathbb{N}$ and $\varepsilon > 0$,

$$P\left(\cup_{\ell=1}^{\infty} A_{\ell,m}(\varepsilon)\right) = \lim_{\ell\to\infty} P(A_{\ell,m}(\varepsilon)) = 1 - \lim_{\ell\to\infty} P(\widetilde{A}_{\ell,m}(\varepsilon)) = 1. \qquad (2.10)$$

Next, denote $N = \cup_{m=1}^{\infty} \cup_{k=1}^{\infty} N_{m,k}$, where $N_{m,k} = \Omega\backslash \cup_{\ell=1}^{\infty} A_{\ell,m}(1/k)$, and note that by (2.10), $P(N_{m,k}) = 0$, hence $P(N) \leq \sum_{m=1}^{\infty}\sum_{k=1}^{\infty} P(N_{m,k}) = 0$. Moreover, note that $\omega \in \Omega\backslash N$ implies that for all $m \in \mathbb{N}$ and $\varepsilon > 0$,

$\omega \in \cup_{\ell=1}^{\infty} A_{\ell,m}(\varepsilon)$, hence $\omega \in \cap_{m=1}^{\infty} \cup_{\ell=1}^{\infty} A_{\ell,m}(\varepsilon)$, which in its turn implies that for all $m \in \mathbb{N}$ there exists an $\ell_0(\varepsilon) \in \mathbb{N}$ such that for all $k \geq \ell_0(\varepsilon)$,

$$|X_{n_k}(\omega) - X_{n_{k+m}}(\omega)| \leq \varepsilon.$$

Consequently, $X_{n_k}(\omega)$ is a Cauchy sequence in $\mathbb{R}$ and therefore by Theorem 2.2, $X_{n_k}(\omega)$ converges to a limit $X(\omega)$ in $\mathbb{R}$, which is measurable $\mathcal{F}$.[6] Thus for $k \to \infty$, $X_{n_k} \stackrel{\text{a.s.}}{\to} X$, where $X$ is a random variable defined on $\{\Omega, \mathcal{F}, P\}$. Hence, for fixed $m$ and $k \to \infty$,

$$(X_{n_k} - X_m)^2 \stackrel{\text{a.s.}}{\to} (X - X_m)^2 . \tag{2.11}$$

Finally, it follows from (2.11), Fatou's lemma[7] 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 $m \to \infty$.

## 2.7.7 Theorem 2.7

Let $\boldsymbol{\delta}_n = \{\delta_{n,m}\}_{m=1}^{\infty}$ be a Cauchy sequence in $\mathbb{R}^{\infty}$. Then for each $m \in \mathbb{N}$, $\delta_{n,m}$ is a Cauchy sequence in $\mathbb{R}$, so that by Theorem 2.2, $\lim_{n\to\infty} \delta_{n,m} = \delta_m \in \mathbb{R}$, hence, for each $L, k \in \mathbb{N}$,

$$\begin{aligned} \sum_{m=1}^{L} \delta_m^2 &= \lim_{n\to\infty} \sum_{m=1}^{L} \delta_{n,m}^2 \leq 2 \limsup_{n\to\infty} \sum_{m=1}^{L} (\delta_{n,m} - \delta_{k,m})^2 + 2 \sum_{m=1}^{L} \delta_{k,m}^2 \\ &\leq 2 \limsup_{n\to\infty} \sum_{m=1}^{\infty} (\delta_{n,m} - \delta_{k,m})^2 + 2 \sum_{m=1}^{\infty} \delta_{k,m}^2. \end{aligned}$$

By the Cauchy property we can choose $k$ so large that $\limsup_{n\to\infty} \sum_{m=1}^{\infty} (\delta_{n,m} - \delta_{k,m})^2 < \varepsilon$, hence $\sum_{m=1}^{L} \delta_m^2 \leq 2\varepsilon + 2 \sum_{m=1}^{\infty} \delta_{k,m}^2$ and therefore, letting $L \to \infty$, we have $\sum_{m=1}^{\infty} \delta_m^2 \leq 2\varepsilon + 2 \sum_{m=1}^{\infty} \delta_{k,m}^2 < \infty$. Thus, $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} \in \mathbb{R}^{\infty}$.

---

[6] 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).

[7] 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).

Finally, it remains to show that $\lim_{n\to\infty}||\boldsymbol{\delta}_n - \boldsymbol{\delta}|| = 0$, as follows. First, note that for $k, L \in \mathbb{N}$,

$$\begin{aligned}\sum_{m=L+1}^{\infty}(\delta_{n,m} - \delta_m)^2 &\leq 2\sum_{m=L+1}^{\infty}(\delta_{n,m} - \delta_{k,m})^2 + 2\sum_{m=L+1}^{\infty}(\delta_{k,m} - \delta_m)^2 \\ &\leq 2\sum_{m=1}^{\infty}(\delta_{n,m} - \delta_{k,m})^2 + 2\sum_{m=L+1}^{\infty}(\delta_{k,m} - \delta_m)^2.\end{aligned}$$

Again, given an arbitrary $\varepsilon > 0$ we can choose $k$ so large that for all $n \geq k$, $\sum_{m=1}^{\infty}(\delta_{n,m} - \delta_{k,m})^2 < \varepsilon/4$. Moreover, since $\sum_{m=1}^{\infty}(\delta_{k,m} - \delta_m)^2 < \infty$, we can choose $L$ so large that $\sum_{m=L+1}^{\infty}(\delta_{k,m} - \delta_m)^2 < \varepsilon/4$. Thus, for these $L$ and $k$, $\limsup_{n\to\infty}\sum_{m=L+1}^{\infty}(\delta_{n,m} - \delta_m)^2 < \varepsilon$, whereas obviously, $\limsup_{n\to\infty}\sum_{m=1}^{L}(\delta_{n,m} - \delta_m)^2 = 0$. Hence, $\limsup_{n\to\infty}||\boldsymbol{\delta}_n - \boldsymbol{\delta}||^2 < \varepsilon$, which implies $\lim_{n\to\infty}||\boldsymbol{\delta}_n - \boldsymbol{\delta}|| = 0$.

# 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 $\beta = \left(X'X\right)^{-1} X'y$, hence $\widehat{y} = X\beta = X\left(X'X\right)^{-1} X'y$.

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 $\beta$ accordingly as $\beta = (b_1', b_2')'$. Then

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

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

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

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 in Theorem 3.1, 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 $\varphi$ 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 $\alpha$, $\beta$, and $\gamma$, 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 $x_1, ..., x_n$: $\mathcal{M}_n = \text{span}(\{x_k\}_{k=1}^{n})$. As we have seen from Lemma 2.2, $\mathcal{M}_n$ 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.

## 3.4 Regularity

**Definition 3.2**. *Denote* $v_1 = x_1$, *and let* $\widehat{z}_{k+1}$ *be the projection of* $x_{k+1}$ *on* $\text{span}(\{x_m\}_{m=1}^{k})$, *with residual* $v_{k+1} = x_{k+1} - \widehat{z}_{k+1}$. *If* $||v_k|| > 0$ *for all* $k \in \mathbb{N}$ *then* $\{x_k\}_{k=1}^{\infty}$ *is said to be left-regular.*

Another form of regularity is the following.

**Definition 3.3**. *Let* $\{x_k\}_{k=1}^{\infty}$ *be a sequence of elements of a Hilbert space* $\mathcal{H}$. *Denote the projection of* $x_k$ *on* $\text{span}(\{x_j\}_{j=k+1}^{\infty})$ *by* $\widehat{x}_k$, *and let* $u_k = x_k - \widehat{x}_k$. *The sequence* $\{x_k\}_{k=1}^{\infty}$ *is said to be right-regular if* $||u_k|| > 0$ *for all* $k \in \mathbb{N}$.

Note that if $||x_1|| = 0$ then $||u_1|| = 0$ because $\text{span}(\{x_j\}_{j=2}^{\infty})$ contains the null element. If $||v_{k+1}|| = 0$ for some $k \in \mathbb{N}$ then $x_{k+1} = \sum_{m=1}^{k} c_m x_m$ for a nonzero vector $(c_1, ..., c_k)' \in \mathbb{R}^k$. Let $j$ be the smallest $m$ for which $c_m \neq 0$, so that $x_j = -\sum_{m=j+1}^{k}(c_m/c_j)x_m + c_j^{-1}x_{k+1}$. Hence, $x_j \in \text{span}(\{x_m\}_{m=j+1}^{k+1} \subset \text{span}(\{x_m\}_{m=j+1}^{\infty}$and thus $||u_j|| = 0$. Consequently, if $\{x_k\}_{k=1}^{\infty}$ is right-regular then it is left-regular as well.

The converse, however, is in general not true: left-regularity does not imply right-regularity. As a counter-example, consider the sequence $x_1 = 1$, $x_2 = \exp(u) - 1$, $x_k = u^{k-2}$ for $k \geq 3$ in $\mathcal{H} = L^2(0,1)$. It is easy to verify that this sequence is left-regular, but since $x_2 = \exp(u) - 1 = \sum_{m=1}^{\infty} u^m/m! \in \text{span}(\{x_k\}_{k=3}^{\infty})$, it is not right-regular.

Let $\{x_k\}_{k=1}^{\infty}$ be left-regular, and denote $e_k = ||v_k||^{-1}v_k$, where $v_k$ is defined in Definition 3.2. Then $\{e_k\}_{k=1}^{\infty}$ is an orthonormal sequence. It is easy to verify that for all $n \in \mathbb{N}$,

$$\text{span}(\{e_k\}_{k=1}^{n}) = \text{span}(\{x_k\}_{k=1}^{n}) \tag{3.3}$$

hence

$$\begin{aligned} \text{span}(\{e_k\}_{k=1}^{\infty}) \stackrel{\text{def.}}{=} \overline{\cup_{n=1}^{\infty}\text{span}(\{e_k\}_{k=1}^{n})} &= \overline{\cup_{n=1}^{\infty}\text{span}(\{x_k\}_{k=1}^{n})} \\ &= \text{span}(\{x_k\}_{k=1}^{\infty}) \end{aligned} \tag{3.4}$$

Similarly, let $\{x_k\}_{k=1}^{\infty}$ be right-regular, and denote $e_k^* = ||u_k||^{-1}u_k$, where $u_k$ is defined in Definition 3.3. Then $\{e_k^*\}_{k=1}^{\infty}$ is also orthonormal, but now $\text{span}(\{e_k^*\}_{k=1}^{\infty}) \subset \text{span}(\{x_k\}_{k=1}^{\infty})$. The latter follows from the Wold decomposition theorem (Theorem 4.1) below.

## 3.5 Convergence of projections

The following convergence results for projections do not require regularity conditions

**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 a counter example,[2] 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 $n \geq 3$,

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

[2] Thanks to Peter Boswijk, Univeristy of Amsterdam, for pointing out an error in a previous version of this counter example.

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

The solutions of these normal equations are

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

hence

$$\widehat{X}_{t,n} = \sum_{j=1}^{n} \left( \frac{j}{n+1} - 1 \right) X_{t-j} \tag{3.5}$$

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

$$\lim_{n\to\infty} \left\| \widehat{X}_t - \sum_{j=1}^{n} \theta_j X_{t-j} \right\|^2 = \lim_{n\to\infty} E\left[ \left( \widehat{X}_t - \sum_{j=1}^{n} \theta_j X_{t-j} \right)^2 \right] = 0. \tag{3.6}$$

If so, it follows from Theorem 3.2 and (3.5) that

$$\begin{aligned}0 &= \lim_{n\to\infty} E\left[ \left( \sum_{j=1}^{n} \theta_j X_{t-j} - \sum_{j=1}^{n} \left( \frac{j}{n+1} - 1 \right) X_{t-j} \right)^2 \right] \qquad (3.7) \\ &= \lim_{n\to\infty} E\left[ \left( \sum_{j=1}^{n} \left( \frac{j}{n+1} - 1 - \theta_j \right) X_{t-j} \right)^2 \right]\end{aligned}$$

However,

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

hence

$$E\left[\left(\sum_{j=1}^{n}\left(\frac{j}{n+1}-1-\theta_j\right)X_{t-j}\right)^2\right]=\left(\frac{n}{n+1}+\theta_1\right)^2$$
$$+\sum_{j=1}^{n-1}\left(\theta_{j+1}-\theta_j-\frac{1}{n+1}\right)^2+\left(\frac{1}{n+1}+\theta_n\right)^2. \qquad (3.8)$$

This equality implies that for arbitrary $m \in \mathbb{N}$,

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

Therefore, a necessary condition for (3.7) is that $\theta_m = -1$ for $m \in \mathbb{N}$. But then it follows from (3.8) 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 (3.7). Thus, in this case there does **not** exist a sequence $\{\theta_j\}_{j=1}^{\infty}$ such that (3.6) holds.

The problem that for the projection $\widehat{z}$ on $\operatorname{span}(\{x_j\}_{j=1}^{\infty})$ 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 $\{x_k\}_{k=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ is orthonormal, i.e.,*

$$\langle x_i, x_k\rangle = I(i=k), \qquad (3.9)$$

*then the projection $\widehat{z}$ of $z \in \mathcal{H}$ on $\operatorname{span}(\{x_k\}_{k=1}^{\infty})$ takes the form $\widehat{z}=\sum_{k=1}^{\infty}\theta_k x_k$ (in the sense that $\lim_{n\to\infty}||\widehat{z}-\sum_{k=1}^{n}\theta_k x_k||=0$), where $\theta_k=\langle z, x_k\rangle$ with $\sum_{k=1}^{\infty}\theta_k^2=||\widehat{z}||^2\leq||z||^2<\infty$.*

**Remark**. The latter inequality is known as the Bessel's inequality. Moreover, if $\{x_k\}_{k=1}^{\infty}$ is complete, i.e., $\text{span}(\{x_k\}_{k=1}^{\infty}) = \mathcal{H}$, then $\widehat{z} = z$, so that $\sum_{k=1}^{\infty} \theta_k^2 = ||z||^2$. This equality is known as Parseval's equality.

Similarly, the following results hold.

**Theorem 3.4.** *Let $\{x_k\}_{k=1}^{\infty}$ be a left-regular sequence in a Hilbert space $\mathcal{H}$, and let $e_k = ||v_k||^{-1} v_k$, where $v_k$ is defined in Definition 3.2. Then the projection $\widehat{z}$ of $z \in \mathcal{H}$ on $\text{span}(\{x_k\}_{k=1}^{\infty})$ takes the form $\widehat{z} = \sum_{k=1}^{\infty} \alpha_k e_k$ (in the sense that $\lim_{n\to\infty} ||\widehat{z} - \sum_{k=1}^{n} \alpha_k x_k|| = 0$), where $\alpha_k = \langle z, e_k \rangle$ with $\sum_{k=1}^{\infty} \alpha_k^2 = ||\widehat{z}||^2 \leq ||z||^2 < \infty$. Consequently, every $z \in \mathcal{H}$ can be written as $z = \sum_{k=1}^{\infty} \alpha_k e_k + u$, where the $\alpha_k$'s are the same as before, with $\langle u, e_k \rangle = 0$ for all $k \in \mathbb{N}$.*[3]

Theorem 3.4 follows straightforwardly from Theorems 3.1 and 3.3 and the equalities (3.3) and (3.4).

## 3.6 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

$$||y - z_n|| \leq ||y - \widehat{y}|| + n^{-1}. \tag{3.10}$$

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

$$\lim_{n\to\infty} ||y - z_n||^2 = ||y - \widehat{y}||^2 = \delta. \tag{3.11}$$

say. The first step is to show that $z_n$ is a Cauchy sequence, as follows. Observe that

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

[3] Note that if $z \in \text{span}(\{x_k\}_{k=1}^{\infty})$ then $u = 0$.

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

$$||z_n - z_m||^2 = 2||z_n - y||^2 + 2||z_m - y||^2 - 4.||0.5(z_n + z_m) - y||^2 \quad (3.12)$$

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 (3.10) and (3.11), $||z_n - y||^2 \leq (\delta + n^{-1})^2$ and $||z_m - y||^2 \leq (\delta + m^{-1})^2$. Therefore, it follows from (3.12) 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

$$\lim_{n\to\infty} ||\widehat{z} - z_n|| = 0. \tag{3.13}$$

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 (3.13) that

$$\lim_{n\to\infty} \delta_n = \delta. \tag{3.14}$$

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

$$\begin{aligned} ||\widehat{z} - \widehat{z}_n||^2 &= ||z - \widehat{z}_n - u||^2 = ||z - \widehat{z}_n||^2 + ||u||^2 - 2\langle z - \widehat{z}_n, u\rangle \\ &= ||z - \widehat{z}_n||^2 + ||u||^2 - 2\langle z, u\rangle = \delta_n^2 - \delta^2, \end{aligned} \tag{3.15}$$

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 (3.14) and (3.15).

### 3.6.4 Theorem 3.3

Due to the orthonormality condition (3.9), the projection $\widehat{z}_n$ of $z$ on $\mathcal{M}_n = \text{span}(\{x_j\}_{j=1}^n)$ takes the form

$$\widehat{z}_n = \sum_{j=1}^n \theta_j x_j, \text{ where } \theta_j = \langle z, x_j\rangle. \tag{3.16}$$

Moreover, denoting $u_n = z - \widehat{z}_n$, it follows from (3.9) and (3.16) that

$$\begin{aligned} ||u_n||^2 &= \left\| z - \sum_{j=1}^n \theta_j x_j \right\|^2 = ||z||^2 - 2\sum_{j=1}^n \theta_j \langle z, x_j\rangle + \sum_{j=1}^n \sum_{i=1}^n \theta_j \theta_i \langle x_j, x_i\rangle \\ &= ||z||^2 - \sum_{j=1}^n \theta_j^2 \geq 0 \end{aligned} \tag{3.17}$$

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 $\widehat{z} = \sum_{j=1}^{\infty} \theta_j x_j$.

# Part II

# The Wold decomposition and its time series applications

# Chapter 4

# The Wold decomposition

The original Wold decomposition, due to Wold (1938), applies to zero-mean covariance stationary time series processes. See Theorem 4.2 below. I will prove the latter theorem on the basis of the following general version of the Wold decomposition..

**Theorem 4.1**. *Given a right-regular sequence $\{x_k\}_{k=1}^{\infty}$ in a Hilbert space $\mathcal{H}$, every $x \in \mathcal{S} = \text{span}(\{x_k\}_{k=1}^{\infty})$ can be written as*

$$x = \sum_{k=1}^{\infty} \alpha_k e_k + w, \tag{4.1}$$

*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*

$$w \in \mathcal{S}_{\infty} \cap \mathcal{U}_{\infty}^{\perp}, \tag{4.2}$$

*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 (4.2) implies that $w$ is orthogonal to all the $e_k$'s: $\langle e_k, w \rangle = 0$ for all $k \in \mathbb{N}$.*

Each $e_k$ in (4.1) is the normalized residual of the projection of $x_k$ on $\text{span}(\{x_m\}_{m=k+1}^{\infty})$, and therefore the orthonormal sequence $\{e_k\}_{k=1}^{\infty}$ depends on $\{x_k\}_{k=1}^{\infty}$ only. However, the term $w$ in (4.1) depends on $x$ because it is the residual of the projection of $x$ on $\text{span}(\{e_k\}_{k=1}^{\infty})$, so that by Theorem 3.1, $\langle e_k, w \rangle = 0$ for all $k \in \mathbb{N}$, hence $w \in \mathcal{U}_{\infty}^{\perp}$. Therefore, the actual contents of Theorem 4.1 is that $w \in \mathcal{S}_{\infty}$.

In general, $w \neq 0$, as will be demonstrated for the next version of the Wold decomposition. Therefore, the main difference between Theorems 3.5 and 4.1 is that in the case of Theorem 3.5 (4.1) holds with $w = 0$.

In the proof of Theorem 4.1 in the appendix to this chapter I will need the following generalization of Definition 2.10.

**Definition 4.1.** *Let $\mathcal{S}_1, ..., \mathcal{S}_n$ be subspaces of a Hilbert space $\mathcal{H}$. Then* $\text{span}(\mathcal{S}_1, ..., \mathcal{S}_n)$ *is the closure of the space of all linear combinations $\sum_{j=1}^{n} c_j x_j$, where $x_j \in \mathcal{S}_j$.*

In the case of the Hilbert space $\mathcal{R}_0$ of zero-mean random variables with finite second moments, inner product $\langle X, Y \rangle = E[X.Y]$ and associated norm and metric, the results of Theorem 4.1 translate as follows:

**Theorem 4.2**. *Let $X_t$ be a regular univariate zero-mean covariance stationary time series process. Then $X_t$ can be written as*

$$X_t = \sum_{j=0}^{\infty} \alpha_j U_{t-j} + W_t \ a.s., \tag{4.3}$$

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

$$\alpha_j = E[X_t U_{t-j}], \ \sum_{j=0}^{\infty} \alpha_j^2 < \infty, \tag{4.4}$$

*and $W_t$ is a zero-mean covariance stationary process satisfying*

$$W_t \in \mathcal{U}_t^{\perp} \cap \mathcal{S}_{-\infty}, \tag{4.5}$$

*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* (4.5) *implies that*

$$W_t \in \ \text{span}\left(\{W_{t-m}\}_{m=1}^{\infty}\right), \tag{4.6}$$

*which in its turn implies that $W_t$ is perfectly predictable from the past values $W_{t-1}, W_{t-2}, W_{t-3}, ........$ Moreover,* (4.5) *implies that*

$$E[W_t U_{t-m}] = 0 \tag{4.7}$$

*for all leads and lags $m$.*

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

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 $E[W_t] = 0$ and $E[W_t W_{t-m}] = \cos(\lambda m)$, 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 $\lambda$, hence $W_t$ is then determined for all $t$.

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 $\langle X, Y \rangle = E[X'Y]$ 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 $\Sigma = E[V_t V_t']$. In this case we need to extend the notion of regularity by requiring that $\Sigma$ is positive definite rather than only $||V_t||^2 = E[V_t' V_t] > 0$, 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 4.1 and 4.2 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 4.2.

The question now arises under which conditions the deterministic process $W_t$ is a.s. equal 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 $\sigma$-algebra of the process $X_t$:

**Definition 3.2.** *Let $\mathcal{F}_t = \sigma(\{X_{t-j}\}_{j=0}^{\infty})$ be the $\sigma$-algebra generated by $\{X_{t-j}\}_{j=0}^{\infty}$. The $\sigma$-algebra $\mathcal{F}_{-\infty} = \cap_t \mathcal{F}_t$ is called the remote $\sigma$-algebra of the process $X_t$.*

If the process $X_t$ is independent then it follows from Kolmogorov's zero-one law[1] 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 $\alpha$-mixing processes.[2] This gives rise to the notion of vanishing memory:

**Definition 3.3.** *A time series process is said to have a vanishing memory if the sets in its remote $\sigma$-algebra $\mathcal{F}_{-\infty}$ have either probability one or zero, i.e., $A \in \mathcal{F}_{-\infty}$ implies that either $P[A] = 1$ or $P[A] = 0$.*

In that case $E[W_t|\mathcal{F}_{-\infty}] = E[W_t]$ a.s.[3] 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 4.3**. *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.*

## 4.1 Applications

### 4.1.1 ARMA models

The Wold decomposition theorem in the form of Theorem 4.2 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, $\alpha\,(L) = 1 + \sum_{k=1}^{\infty} \alpha_k L^k$, and the $U_t$'s are zero-mean uncorrelated covariance stationary random variables. The function $\alpha\,(L)$ can be approximated arbitrarily close by a ratio of two lag polynomials,

[1] See for example Bierens (2004, Theorem 7.5, p.185).
[2] See for example Bierens (2004, Theorem 7.6, p.186).
[3] 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[4] with inverse $\varphi_p^{-1}(L)$. 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$(\infty)$ representation

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

This representation plays a key role in forecasting.

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

### 4.1.2 Innovation response analysis

An important econometric application of the multivariate version of the Wold decomposition is Sims' (1980) innovation response analysis. Sims' (1980) landmark paper[5] 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:

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

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

[5] For which he was awarded the 2011 Nobel Prize in Economics, jointly with Thomas Sargent.

Assume that the error vectors $U_t$ are i.i.d. $N_k[0, \Sigma]$, where $\Sigma$ 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 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 $\Sigma$, $\mu$, 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 $\Sigma$ of the $U_t$'s can be written as $\Sigma = \Delta.\Delta'$, where $\Delta$ 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 $\delta_i$ is column $i$ of $\Delta$, 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.

## 4.2 Proofs

### 4.2.1 Theorem 4.1

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

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

[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 $|z| > 1$.

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

$$\mathcal{U}_{n+1}^\perp \subset \mathcal{U}_n^\perp. \tag{4.8}$$

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

$$\mathcal{M}_n \subset \text{ span}\left(\mathcal{U}_n, \mathcal{U}_n^\perp \cap \mathcal{S}_2\right). \tag{4.9}$$

**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

$$\sum_{k=1}^n c_k \widehat{x}_k \in \mathcal{S}_2 \tag{4.10}$$

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 (4.10). 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, (4.9) holds.

**Step 2**. I will now show that

$$\text{span}\left(\mathcal{U}_n, \mathcal{U}_n^{\perp} \cap \mathcal{S}_2\right) = \text{span}\left(\mathcal{U}_n, \mathcal{U}_n^{\perp} \cap \mathcal{S}_{n+1}\right). \tag{4.11}$$

**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 exist 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,

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

because $\sum_{k=2}^{n} c_k \widehat{x}_k \in \mathcal{S}_3$, $\sum_{k=n+1}^{m} c_k x_k \in \mathcal{S}_{n+1}$, and $e_2$ is orthogonal to $\mathcal{S}_3$ and $\mathcal{S}_{n+1}$. 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

$$\mathcal{U}_n^{\perp} \cap \left(\cup_{m=n+1}^{\infty} \mathcal{S}_{2,m}\right) = \mathcal{U}_n^{\perp} \cap \left(\cup_{m=n+1}^{\infty} \mathcal{S}_{n+1,m}\right) \tag{4.12}$$

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

$$\mathcal{U}_n^{\perp} \cap \left(\cup_{m=2}^{\infty} \mathcal{S}_{2,m}\right) = \mathcal{U}_n^{\perp} \cap \left(\cup_{m=n+1}^{\infty} \mathcal{S}_{2,m}\right). \tag{4.13}$$

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 (4.13) 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 (4.11) holds.

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

$$\mathcal{S}_1 = \overline{\cup_{n=1}^{\infty} \mathcal{R}_n}. \tag{4.14}$$

**Proof**. Combining (4.9) and (4.11) yields $\mathcal{M}_n \subset \mathcal{R}_n$, hence

$$\mathcal{S}_1 = \overline{\cup_{n=1}^{\infty} \mathcal{M}_n} \subset \overline{\cup_{n=1}^{\infty} \mathcal{R}_n}, \tag{4.15}$$

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

$$\overline{\cup_{n=1}^{\infty} \mathcal{R}_n} \subset \mathcal{S}_1. \tag{4.16}$$

Thus, the result (4.14) follows from (4.15) and (4.16).

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

$$\widehat{x}_n = \sum_{j=1}^{n} \alpha_j e_j + w_{n+1} \tag{4.17}$$

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,

$$\sum_{j=1}^{\infty} \alpha_j^2 < \infty. \tag{4.18}$$

Furthermore,

$$\lim_{n\to\infty} \left\| x - \sum_{j=1}^{n} \alpha_j e_j - w_{n+1} \right\| = 0. \tag{4.19}$$

**Proof**. By the definition of $\mathcal{R}_n$ and by Definition 4.1, $\widehat{x}_n = \sum_{j=1}^{n} \theta_j e_j + w$ for some constants $\theta_j$ and a $w \in \mathcal{U}_n^{\perp} \cap S_{n+1}$. To determine the $\theta_j$'s and $w$, note that

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

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

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

Thus

$$\begin{aligned} ||x - \widehat{x}_n||^2 &= \inf_{\theta_1,\ldots,\theta_n, w \in \mathcal{U}_n^\perp \cap S_{n+1}} \left\| x - \sum_{j=1}^n \theta_j e_j - w \right\|^2 \\ &= \inf_{\theta_1,\ldots,\theta_n, w \in \mathcal{U}_n^\perp \cap S_{n+1}} \left( ||x - w||^2 - 2\sum_{j=1}^n \theta_j \langle e_j, x\rangle + \sum_{j=1}^n \theta_j^2 \right) \\ &= \inf_{w \in \mathcal{U}_n^\perp \cap S_{n+1}} ||x - w||^2 - \sum_{j=1}^n \alpha_j^2 \\ &= ||x - w_{n+1}||^2 - \sum_{j=1}^n \alpha_j^2 \qquad (4.20) \end{aligned}$$

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 $n$,

$$\sum_{j=1}^n \alpha_j^2 \le ||x - w_{n+1}||^2 \le ||x||^2 \qquad (4.21)$$

so that (4.18) holds.

Finally, to prove (4.19), 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 (4.14) 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

$$\lim_{n\to\infty} \|z - z_n\| = 0, \text{ where } z \in \mathcal{U}_\infty. \qquad (4.22)$$

**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

$$\|z_n - z_m\|^2 = \left\| \sum_{j=\min(m,n)+1}^{\max(m,n)} \alpha_j e_j \right\|^2$$

$$\begin{aligned} &= \sum_{j=\min(m,n)+1}^{\max(m,n)} \alpha_j^2 \leq \sum_{j=\min(m,n)+1}^{\infty} \alpha_j^2 \\ &\to 0 \end{aligned}$$

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

$$\lim_{n\to\infty} ||w_{n+1} - w|| = 0. \tag{4.23}$$

**Proof**. Recall from Step 4 that

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

Moreover, it follows from (4.8) and the definition of $S_{n+1}$ that for an arbitrary $k \geq 1$,

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

hence

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

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

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

as $\min(m,n) \to \infty$. Thus, there exists a $w \in \mathcal{U}_k^\perp \cap S_{k+1}$ such that (4.23) 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 (4.18), (4.22), (4.23) 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 $k \in \mathbb{N}$.

### 4.2.2 Theorem 4.2

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 span$(\{X_{t-j}\}_{j=1}^{\infty})$. The uncorrelatedness of the $\widetilde{U}_t$'s follows from Theorem 4.1, 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 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

$$\lim_{n\to\infty} \left\| \widehat{X}_{t,n} - \widehat{X}_t \right\|^2 = \lim_{n\to\infty} E\left[\left(\widehat{X}_{t,n} - \widehat{X}_t\right)^2\right] = 0 \tag{4.24}$$

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

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

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 (4.24) that

$$\lim_{n\to\infty} E\left[\left(\widetilde{U}_t - \widetilde{U}_{t,n}\right)^2\right] = \lim_{n\to\infty} E\left[\left(\widehat{X}_{t,n} - \widehat{X}_t\right)^2\right] = 0. \tag{4.25}$$

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 $\sigma_n^2$ 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 (4.4), (4.5) and (4.7)**

The result of Theorem 4.1 can now be translated as

$$\lim_{n\to\infty}\left\|X_t-\sum_{j=0}^{n}\alpha_jU_{t-j}-W_t\right\|=0, \tag{4.26}$$

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

We still need to prove that the $\alpha_k$'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 (4.25),

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

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 (4.5) and (4.7) follow straightforwardly from Theorem 4.1.

**Proof of (4.3)**

The result (4.26) implies, by Chebyshev's inequality, that

$$X_t=\operatorname*{plim}_{n\to\infty}\sum_{j=0}^{n}\alpha_jU_{t-j}+W_t. \tag{4.27}$$

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 $k_m$,

$$\sum_{j=0}^{k_m}\alpha_jU_{t-j}\overset{a.s.}{\to}X_t-W_t \tag{4.28}$$

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

$$k_{m_{n-1}} < n \leq k_{m_n}. \tag{4.29}$$

Moreover, (4.28) implies that

$$\sum_{j=0}^{k_{m_n}} \alpha_j U_{t-j} \stackrel{\text{a.s.}}{\to} X_t - W_t \text{ as } n \to \infty. \tag{4.30}$$

Due to (4.29),

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

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

$$\sum_{j=0}^{k_{m_n}} \alpha_j U_{t-j} - \sum_{j=0}^{n} \alpha_j U_{t-j} \stackrel{\text{a.s.}}{\to} 0 \text{ as } n \to \infty. \tag{4.31}$$

Combining (4.30) and (4.31) it follow now that

$$\sum_{j=0}^{n} \alpha_j U_{t-j} \stackrel{\text{a.s.}}{\to} X_t - W_t \text{ as } n \to \infty. \tag{4.32}$$

Since $\sum_{j=0}^{\infty} \alpha_j U_{t-j}$ is defined as $\lim_{n\to\infty} \sum_{j=0}^{n} \alpha_j U_{t-j}$, (4.3) is equivalent to (4.32).

[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 (4.3) that $E[W_t] = 0$. Moreover, it is left as an exercise to show that for $m \geq 0$,

$$E[W_t W_{t-m}] = \gamma(m) - \sum_{j=0}^{\infty} \alpha_{j+m}\alpha_j. \tag{4.33}$$

**Proof of (4.6)**
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 (4.3),

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

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 (4.6). ■

# Part III

# Series expansions of functions

# Chapter 5

# Orthogonal polynomials

## 5.1 Introduction

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

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

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

$$p_k(x|w) = \sum_{j=0}^{k} \alpha_{k,j} x^j, \;\; \alpha_{k,k} = 1, \;\; k \in \mathbb{N}_0 \tag{5.1}$$

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

$$\int_{-\infty}^{\infty} p_k(x|w) p_m(x|w) w(x)\mathrm{d}x = 0 \text{ if } k \neq m. \tag{5.2}$$

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

Defining

$$\overline{p}_k(x|w) = \frac{p_k(x|w)}{\sqrt{\int_{-\infty}^{\infty} p_k(y|w)^2 w(y) dy}} \tag{5.3}$$

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

$$\int_{-\infty}^{\infty} \overline{p}_k(x|w)\overline{p}_m(x|w)w(x)\mathrm{d}x = \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1 & \text{if} \quad k = m. \end{cases} \tag{5.4}$$

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 $j < k$,

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

and

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

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,

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

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

$$\int_a^b |x|^k w(x)\mathrm{d}x < \infty \tag{5.5}$$

*for* $k \in \mathbb{N}$. *Denote by* $L^2(w)$ *be the Hilbert space of Borel measurable real functions* $f$ *on* $(a,b)$ *satisfying* $\int_a^b f(x)^2 w(x)\mathrm{d}x < \infty$, *with inner product* $\langle f,g\rangle = \int_a^b f(x)g(x)w(x)\mathrm{d}x$ *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*

$$\gamma_k = \langle f, \overline{p}_k\rangle = \int_a^b f(x)\overline{p}_k(x|w)w(x)\mathrm{d}x. \tag{5.6}$$

*Then*

$$\lim_{n\to\infty} \|f - f_n\| = 0, \tag{5.7}$$

*so that the orthonormal polynomials* $\overline{p}_k(x|w)$ *generated by* $w$ *form a complete orthonormal sequence in* $L^2(w)$. *Moreover, the result* (5.7) *implies that every function* $f \in L^2(w)$ *can be written as*

$$f(x) = \sum_{k=0}^{\infty} \gamma_k \overline{p}_k(x|w) \text{ a.e. on } (a,b). \tag{5.8}$$

The result (5.8) follows from the following more general lemma.

**Lemma 5.1**. *Let* $\{\rho_k(x)\}_{k=0}^{\infty}$ *be a complete orthonormal sequence in the Hilbert space* $L^2(w)$, *where* $w(x)$ *is a possibly multivariate density with support an open subset* $\Xi$ *of a Euclidean space.*[1] *Then for any function* $f \in L^2(w)$,

$$f(x) = \sum_{k=0}^{\infty} \gamma_k \rho_k(x) \text{ a.e. on } \Xi,$$

*where for* $k \in \mathbb{N}_0$, $\gamma_k = \int_\Xi \rho_k(x) f(x) w(x)\mathrm{d}x$.

Note that condition (5.5) holds trivially if the support $(a,b)$ of $w(x)$ is bounded. However, as is well-known, condition (5.5) also holds for the standard normal density, the exponential density and the density of the Gamma distribution, for example.

[1] Thus, $L^2(w)$ is endowed with the inner product $\langle f_1, f_2\rangle = \int_\Xi f_1(x)f_2(x)w(x)\mathrm{d}x$ and associated norm and metric.

Since for every density $w(x)$ with support $(a,b)$, $\int_a^b f(x)^2 \mathrm{d}x < \infty$ implies that $f(x)/\sqrt{w(x)} \in L^2(w)$, the following corollary of Theorem 5.1 holds trivially.

**Corollary 5.1**. *Let $L^2(a,b)$, $-\infty \leq a < b \leq \infty$, be the Hilbert space of square integrable Borel measurable real functions on $(a,b)$, with inner product $\langle f,g\rangle = \int_a^b f(x)g(x)\mathrm{d}x$ 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 (5.5), the $\overline{p}_k(x|w)$'s are the orthonormal polynomials generated by $w(x)$ and the $\gamma_k$'s are the Fourier coefficients of $f(x)/\sqrt{w(x)}$, i.e., $\gamma_k = \int_a^b f(x)\overline{p}_k(x|w)\sqrt{w(x)}\mathrm{d}x$. This result implies that the functions $\psi_k(x|w) = \overline{p}_k(x|w)\sqrt{w(x)}$, $k \in \mathbb{N}$, form a complete orthonormal sequence in $L^2(a,b)$:*

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

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

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

$$G(x) = a + (b-a)W(x)$$

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

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

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

$$(b-a)\int_a^b f(G(x))^2 w(x)\mathrm{d}x = \int_a^b f(G(x))^2 \mathrm{d}G(x) = \int_a^b f(x)^2 \mathrm{d}x < \infty.$$

Hence $f(G(x)) \in L^2(w)$ and thus by Theorem 5.1,

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

where

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

Consequently

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

Note that $\mathrm{d}G^{-1}\left(x\right)/\mathrm{d}x = \mathrm{d}G^{-1}\left(x\right)/\mathrm{d}G\left(G^{-1}\left(x\right)\right) = 1/G'\left(G^{-1}\left(x\right)\right)$, so that

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

Thus,

**Corollary 5.2**. *Let $w$ be a density with bounded support $(a,b)$, satisfying the moment conditions* (5.5). *Let $W$ be the c.d.f. of $w$, with inverse $W^{-1}$. Then the functions*

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

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

## 5.2 The three-term recurrence relation

It follows from (5.1) that $p_0(x|w) \equiv 1$, and it follows from (5.2) that $p_1(x|w) = \alpha_{1,0} + x$ can be constructed by solving $\int_{-\infty}^{\infty} (\alpha_{1,0} + x)\, w(x) \mathrm{d}x = 0$. Hence, given that $w(x)$ is a density, $\alpha_{1,0} = -\int_{-\infty}^{\infty} x.w(x) \mathrm{d}x$. The question now arises how to construct these orthogonal polynomials further for $k \geq 2$.

The answer is the following.

**Theorem 5.2.** *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* (5.2), *with $w(x)$ satisfying the moment conditions* (5.5), *can be generated recursively by the three-term recurrence relation* (*hereafter referred to as TTRR*)

$$p_{k+1}(x|w) + (b_k - x)\, p_k(x|w) + c_k p_{k-1}(x|w) = 0, \;\; k \in \mathbb{N}, \tag{5.9}$$

*where*

$$b_k = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)^2 w(x) \mathrm{d}x}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x) \mathrm{d}x} \tag{5.10}$$

*and*

$$c_k = \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x) \mathrm{d}x}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x) \mathrm{d}x} \tag{5.11}$$

Moreover,

**Theorem 5.3.** *Every sequence $\overline{p}_k(x|w)$ of orthonormal polynomials with respect to a density function $w(x)$ satisfying the moment conditions* (5.5) *can be generated recursively by the TTRR*

$$a_{k+1}.\overline{p}_{k+1}(x|w) + (b_k - x)\, \overline{p}_k(x|w) + a_k.\overline{p}_{k-1}(x|w) = 0, \quad k \in \mathbb{N}, \tag{5.12}$$

*where*

$$a_k = \left| \lim_{|x| \to \infty} \frac{x.\overline{p}_{k-1}(x|w)}{\overline{p}_k(x|w)} \right| \tag{5.13}$$

*and*

$$b_k = \int_{-\infty}^{\infty} x.\overline{p}_k(x|w)^2 w(x) \mathrm{d}x. \tag{5.14}$$

## 5.3 Examples of orthonormal polynomials

### 5.3.1 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 (5.12). These polynomials are known as Hermite[2] polynomials.

It follows from Theorem 5.1 that the Hermite polynomials span the Hilbert space $L^2(w_{\mathcal{N}[0,1]})$, and it follows from Corollary 5.1 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

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

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

### 5.3.2 Laguerre polynomials

The standard exponential density function

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

gives rise to the orthonormal Laguerre[3] 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 $b_k = 2k + 1$.

[2]Charles Hermite (1822-1901).

[3]Edmund Nicolas Laguerre (1834-1886)

Since the moment conditions (5.5) hold for $w_{\mathrm{Exp}}(x)$, it follows from Theorem 5.1 that any Borel measurable real function $f(x)$ satisfying $\int_0^\infty \exp(-x)$ $f(x)^2\mathrm{d}x < \infty$ 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 $\gamma_k = \int_0^\infty \exp(-x)\overline{p}_{k+1}(x|w)f(x)\mathrm{d}x$.

Again, it follows from Corollary 5.1 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

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

### 5.3.3 Legendre polynomials

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

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

generates the orthonormal Legendre[4] polynomials on $[-1,1]$, with TTRR

$$\begin{aligned}\frac{k+1}{\sqrt{2k+3}\sqrt{2k+1}}&\overline{p}_{k+1}(x|w_{\mathcal{U}[-1,1]}) - x.\overline{p}_k(x|w_{\mathcal{U}[-1,1]}) \\ &+ \frac{k}{\sqrt{2k+1}\sqrt{2k-1}}\overline{p}_{k-1}(x|w_{\mathcal{U}[-1,1]}) = 0,\end{aligned} \tag{5.15}$$

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

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

[4] Adrien-Marie Legendre (1752-1833)

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 (5.15) yields the TTRR

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

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

Again, it follows from Theorem 5.1 that any Borel measurable real function $f(u)$ on $[0,1]$ can be written as $f(u) = \sum_{k=0}^{\infty}\gamma_k\overline{p}_k(u|w_{\mathcal{U}[0,1]})$ a.e., where $\gamma_k = \int_0^1 f\left(x\right)\overline{p}_k(u|w_{\mathcal{U}[0,1]})\mathrm{d}u$, hence $L^2(0,1) = \mathrm{span}\left(\left\{\overline{p}_k(u|w_{\mathcal{U}[0,1]}\right\}_{k=0}^{\infty}\right)$.

These shifted Legendre polynomials have been used by Bierens (2008), Bierens and Carvalho (2007) and Bierens and Song (2012) to model semi-nonparametrically the unobserved heterogeneity distribution of interval-censored mixed proportional hazard models and bivariate mixed proportional hazard models, and the value distribution in first-price auction models, respectively.

### 5.3.4 Chebyshev polynomials

#### Chebyshev polynomials on $[-1,1]$

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

$$w_{\mathcal{C}[-1,1]}(x) = \frac{1}{\pi\sqrt{1-x^2}}I\left(|x| < 1\right). \tag{5.16}$$

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

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

hence

$$\frac{\mathrm{d}\arccos(x)}{\mathrm{d}x} = \frac{-1}{\sqrt{1-x^2}} \tag{5.17}$$

Then

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

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

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

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

$$p_{k+1}(x|w_{\mathcal{C}[-1,1]}) - 2xp_k(x|w_{\mathcal{C}[-1,1]}) + p_{k-1}(x|w_{\mathcal{C}[-1,1]}) = 0,\ \ k \in \mathbb{N}, \tag{5.18}$$

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

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

An important practical difference with the other polynomials discussed so far is that Chebyshev polynomials have the closed form[5]:

$$p_k(x|w_{\mathcal{C}[-1,1]}) = \cos\left(k.\arccos(x)\right). \tag{5.19}$$

To see this, observe from (5.17) and the well-known sine-cosine formulas that

$$\int_{-1}^{1} \frac{\cos\left(k.\arccos(x)\right)\cos\left(m.\arccos(x)\right)}{\pi\sqrt{1-x^2}}\mathrm{d}x$$

[5]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.

$$
\begin{aligned}
&= -\frac{1}{\pi}\int_{-1}^{1} \cos\left(k.\arccos(x)\right)\cos\left(m.\arccos(x)\right)\mathrm{d}\arccos(x) \\
&= \frac{1}{\pi}\int_{0}^{\pi} \cos\left(k.\theta\right)\cos\left(m.\theta\right)\mathrm{d}\theta \\
&= \frac{1}{2\pi}\int_{0}^{\pi} \cos\left(\left(k+m\right)\theta\right)\mathrm{d}\theta + \frac{1}{2\pi}\int_{0}^{\pi} \cos\left(\left(k-m\right)\theta\right)\mathrm{d}\theta \\
&= \frac{1}{2}\left(\frac{\sin\left(\left(k+m\right)\pi\right)}{\left(k+m\right)\pi} + \frac{\sin\left(\left(k-m\right)\pi\right)}{\left(k-m\right)\pi}\right) \\
&= \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1/2 & \text{if} \quad k = m > 0, \\ 1 & \text{if} \quad k = m = 0. \end{cases} \qquad (5.20)
\end{aligned}
$$

Moreover, the TTRR (5.18) 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 (5.19) satisfy the TTRR (5.18) and are therefore genuine polynomials.

In view of (5.20) we can now define the orthonormal Chebyshev polynomials as

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

It is trivial to verify that the density (5.16) satisfies the moment condition (5.5), 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 $\overline{p}_n(x|w_{\mathcal{C}[-1,1]})$ 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 5.2**. *For* $j_1, j_2 = 0, 1, 2, ..., n-1$,

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

Now interpret $k$ in Lemma 5.2 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 5.2 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

$$g(t) = \sum_{j=0}^{n-1} c_{j,n} P_{j,n}(t), \text{ where } c_{j,n} = \frac{1}{n}\sum_{k=1}^{n} g(k) P_{j,n}(k).$$

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

$$g(t) \approx \sum_{j=0}^{m} c_{j,n} P_{j,n}(t)$$

for modest values of $m < n-1$. 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 (5.16) yields

$$w_{\mathcal{C}[0,1]}(u) = \frac{2}{\pi\sqrt{1-(2u-1)^2}} = \frac{1}{\pi\sqrt{u(1-u)}}. \qquad (5.21)$$

with corresponding distribution function

$$W_{\mathcal{C}[0,1]}(u) = 1 - \pi^{-1} \arccos(2u - 1), \tag{5.22}$$

and shifted Chebyshev polynomials

$$\overline{p}_k(u|w_{\mathcal{C}[0,1]}) = \begin{cases} 1 & \text{for} \quad k = 0, \\ \sqrt{2}\cos\left(k.\arccos(2u-1)\right) & \text{for} \quad k \in \mathbb{N}. \end{cases} \tag{5.23}$$

Again, it follows from Corollary 5.1 that the orthonormal sequence

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

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

$$f(u) = \sum_{k=0}^{\infty} \gamma_k \psi_k(u) \text{ a.e. on } (0,1), \tag{5.24}$$

where $\gamma_k = \int_0^1 f(u)\,\psi_k(u)\mathrm{d}u$.

## 5.4 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 $\infty \le a_i < b_i \le \infty$, $i = 1, 2$, satisfying the conditions of Theorem 5.1. Consider the space $L^2(w_1 \times w_2)$ of bivariate Borel measurable real functions $f(x,y)$ satisfying

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} w_1(x)w_2(y)f(x,y)^2\mathrm{d}x\mathrm{d}y < \infty, \tag{5.25}$$

endowed with the inner product

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

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

$$\int_{a_1}^{b_1} w_1(x)f(x,y)^2\mathrm{d}x < \infty, \tag{5.26}$$

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

$$f(x,y) = \sum_{k=0}^{\infty} \gamma_k(y)\overline{p}_k(x|w_1) \text{ a.e. on } (a_1,b_1). \tag{5.27}$$

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

Note that by the Cauchy-Schwarz inequality and (5.25),

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

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

$$\gamma_k(y) = \sum_{m=0}^{\infty} \gamma_{k,m}\overline{p}_m(y|w_2) \text{ a.e. on } (a_2,b_2), \tag{5.28}$$

where

$$\begin{aligned}\gamma_{k,m} &= \int_{a_2}^{b_2} w_2(y)\gamma_k(y)\overline{p}_m(y|w_2)\mathrm{d}y \\ &= \int_{a_1}^{b_1}\int_{a_2}^{b_2} w_1(x)w_2(y)f(x,y)\overline{p}_k(x|w_1)\overline{p}_m(y|w_2)\mathrm{d}x\mathrm{d}y \qquad (5.29)\end{aligned}$$

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

Moreover, note that due to (5.25) the restriction (5.26) 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 (5.29) 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

$$\begin{aligned} f(x,y) &= \frac{\exp\left(-\frac{1}{2}\left(x^2+y^2\right)\right)}{2\pi} \\ &\quad \times \left(\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\gamma_{k,m}\overline{p}_k(x|w_{\mathcal{N}[0,1]})\overline{p}_m(y|w_{\mathcal{N}[0,1]})\right)^2 \qquad (5.30) \\ &\quad \text{a.e. on } \mathbb{R}^2, \text{ where } \sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\gamma_{k,m}^2 = 1. \end{aligned}$$

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 (5.30) of the error density.

## 5.5 Proofs

### 5.5.1 Theorem 5.1

Let $\overline{f}_n(x) = \sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w)$, where $\gamma_k = \int_a^b \overline{p}_k(x|w)f(x)w(x)\mathrm{d}x$, and observe that due to condition (5.5), $\overline{f}_n \in L^2(w)$. Next, observe that

$$\begin{aligned} \left\|f-\overline{f}_n\right\|^2 &= \int_a^b \left(f(x) - \sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w)\right)^2 w(x)\mathrm{d}x \\ &= \int_a^b f(x)^2 w(x)\mathrm{d}x - 2\sum_{k=0}^{n}\gamma_k\int_a^b \overline{p}_k(x|w)f(x)w(x)\mathrm{d}x \\ &\quad + \sum_{k_1=0}^{n}\sum_{k_1=0}^{n}\gamma_{k_1}\gamma_{k_2}\int_a^b \overline{p}_{k_1}(x|w)\overline{p}_{k_2}(x|w)w(x)\mathrm{d}x \\ &= \int_a^b f(x)^2 w(x)\mathrm{d}x - \sum_{k=0}^{n}\gamma_k^2 \geq 0. \qquad (5.31) \end{aligned}$$

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

$$\sum_{k=0}^{\infty}\gamma_k^2 < \infty. \tag{5.32}$$

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

$$\lim_{n\to\infty}\left\|\overline{f}_n - \overline{f}\right\| = 0. \tag{5.33}$$

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

$$\overline{f}(x) = \sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w) + \varepsilon_n(x) \tag{5.34}$$

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

$$\lim_{n\to\infty}\int_a^b \varepsilon_n(x)^2 w(x)\mathrm{d}x = 0. \tag{5.35}$$

**Proof of (5.7)**

To prove (5.7), it suffices to show that

$$\int_a^b \exp(\mathbf{i}.t.x)\left(f(x) - \overline{f}(x)\right)w(x)\mathrm{d}x = 0 \tag{5.36}$$

for all $t \in \mathbb{R}$, because (5.36) implies that $f(x) = \overline{f}(x)$ a.e. on $(a,b)$, due to the uniqueness of the Fourier transform.[6]

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

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

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

hence by (5.35),

$$\int_a^b \left(f(x) - \overline{f}(x)\right) \overline{p}_m(x|w)w(x)\mathrm{d}x = 0 \tag{5.37}$$

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

$$\int_a^b \left(f(x) - \overline{f}(x)\right) x^m w(x)\mathrm{d}x = 0 \text{ for all } m \in \mathbb{N}. \tag{5.38}$$

In its turn (5.38) 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 $n \in \mathbb{N}$,

$$\begin{aligned}
&\int_a^b \exp(\mathbf{i}.t.x) \left(f(x) - \overline{f}(x)\right) w(x)\mathrm{d}x \\
&\quad = \int_a^b \sum_{m=0}^{n} \frac{(\mathbf{i}.t.x)^m}{m!} \left(f(x) - \overline{f}(x)\right) w(x)\mathrm{d}x \\
&\quad + \int_a^b \left(\sum_{m=n+1}^{\infty} \frac{(\mathbf{i}.t.x)^m}{m!}\right) \left(f(x) - \overline{f}(x)\right) w(x)\mathrm{d}x \\
&\quad = \int_a^b \left(\sum_{m=n+1}^{\infty} \frac{(\mathbf{i}.t.x)^m}{m!}\right) \left(f(x) - \overline{f}(x)\right) w(x)\mathrm{d}x
\end{aligned}$$

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

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

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

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

whereas by dominated convergence

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

Since $\varepsilon > 0$ is arbitrary, we therefore have in either case that (5.36) holds. It follows now from (5.34) and (5.35) that

$$\lim_{n\to\infty}\int_a^b \left(f(x) - \sum_{k=0}^{n}\gamma_k \overline{p}_k(x|w)\right)^2 w(x)\,\mathrm{d}x = 0. \tag{5.39}$$

This completes the proof of (5.7).

### 5.5.2 Lemma 5.1

Denote $f_n(x) = \sum_{k=0}^{n}\gamma_k\rho_k(x)$. By the completeness of the orthonormal sequence $\{\rho_k(x)\}_{k=0}^{\infty}$ in $L^2(w)$ it follows that

$$\lim_{n\to\infty}\int_{\Xi}\left(f(x)-f_n(x)\right)^2 w(x)\mathrm{d}x = 0.$$

Now let $X$ be a random drawing from $w(x)$, so that

$$\lim_{n\to\infty} E\left[\left(f(X)-f_n(X)\right)^2\right] = 0.$$

Then by Chebyshev's inequality,

$$f(X) = \operatorname*{plim}_{n\to\infty}\sum_{k=0}^{n}\gamma_k\rho_k(X). \tag{5.40}$$

As is well-known[7], convergence in probability is equivalent to almost sure (a.s.) convergence along a further subsequence of an arbitrary subsequence

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

of $n$. Thus it follows from (5.40) that for any subsequence $n_j$ in $\mathbb{N}$ there exists a further subsequence $n_{j_m}$ such that for $m \to \infty$,

$$\sum_{k=0}^{n_{j_m}} \gamma_k \rho_k(X) \stackrel{\text{a.s.}}{\to} f(X). \tag{5.41}$$

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 $n \to \infty$,

$$\sum_{k=0}^{j_n} \gamma_k \rho_k(X) \stackrel{\text{a.s.}}{\to} f(X). \tag{5.42}$$

The latter implies that

$$\begin{aligned} E\left[\left(\sum_{k=0}^{j_n} \gamma_k \rho_k(X) - \sum_{k=0}^{n} \gamma_k \rho_k(X)\right)^2\right] &= E\left(\sum_{k=n+1}^{j_n} \gamma_k \rho_k(X)\right)^2 \\ &\leq \sum_{k=j_{n-1}+1}^{j_n} \gamma_k^2 \end{aligned}$$

so that

$$\begin{aligned} \sum_{n=1}^{\infty} E\left[\left(\sum_{k=0}^{j_n} \gamma_k \rho_k(X) - \sum_{k=0}^{n} \gamma_k \rho_k(X)\right)^2\right] &\leq \sum_{n=1}^{\infty} \sum_{k=j_{n-1}+1}^{j_n} \gamma_k^2 \\ &\leq \sum_{k=0}^{\infty} \gamma_k^2 < \infty. \end{aligned}$$

Then by Chebyshev's inequality,

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

for all $\varepsilon > 0$, which by the Borel-Cantelli lemma[8] implies that for $n \to \infty$,

$$\sum_{k=0}^{j_n} \gamma_k \rho_k(X) - \sum_{k=0}^{n} \gamma_k \rho_k(X) \stackrel{\text{a.s.}}{\to} 0. \tag{5.43}$$

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

Combining (5.42) and (5.43), it follows that $\sum_{k=0}^{n}\gamma_k\rho_k(X) \overset{\text{a.s.}}{\to} f(X)$ as $n \to \infty$, which implies that $\sum_{k=0}^{n}\gamma_k\rho_k(x) \overset{\text{a.e.}}{\to} f(x)$ on $\Xi$ because the support of $w(x)$ was assumed to be $\Xi$.

### 5.5.3 Theorem 5.2

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 $p_0(x|w), p_1(x|w),...,p_k(x|w)$:

$$p_{k+1}(x|w) - x.p_k(x|w) = \sum_{j=0}^{k}\delta_{j,k}p_j(x|w) \tag{5.44}$$

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

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

so that

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

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 (5.44) that

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

where

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

and

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

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

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

### 5.5.4 Theorem 5.3

Let $d_k = \sqrt{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)\mathrm{d}x}$, 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 (5.9), (5.10) and (5.11) yields

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

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

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

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 5.2.

### 5.5.5 Lemma 5.2

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

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

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

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

$$-\frac{\exp(-\mathbf{i}.\pi m/(2n))}{\exp(\mathbf{i}.\pi m/(2n))-\exp(-\mathbf{i}.\pi m/(2n))}(\cos(\pi m)-1)$$
$$=\cos(\pi m)-1$$

and similarly for $m \neq 0$,

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

Thus, for $j_1 \neq j_2$,

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

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

$$\sum_{k=1}^{n}p_{j}(x_{n,k}|w_{\mathcal{C}[-1,1]})p_{j}(x_{n,k}|w_{\mathcal{C}[-1,1]})$$

$$= \frac{1}{2} \cos(\pi.j/n) \sum_{k=1}^{n} \cos(2\pi.j.k/n) + \frac{1}{2} \sin(\pi.j/n) \sum_{k=1}^{n} \sin(2\pi.j.k/n)$$

$$+\frac{1}{2}n = \frac{1}{2}n$$

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

# Chapter 6

# Trigonometric series

## 6.1 Cosine series representation

Note that the distribution function $W_{\mathcal{C}[0,1]}(u)$ defined by (5.22) has inverse

$$W_{\mathcal{C}[0,1]}^{-1}(u) = (1 + \cos(\pi(1-u)))/2. \tag{6.1}$$

It is now easy to verify from Corollary 5.2, (6.1) and (5.24) that every function $f \in L^2(0,1)$ can be written as

$$\begin{aligned} f(u) &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2} \cos\left(k.\arccos(2W_c^{-1}(u) - 1)\right) \qquad (6.2) \\ &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2} \cos(k\pi(1-u)) \\ &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k (-1)^k \sqrt{2} \cos(k\pi u) \\ &= \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos(k\pi u) \ \text{ a.e. on } (0,1), \end{aligned}$$

where

$$\begin{aligned} \alpha_k &= (-1)^k \gamma_k \\ &= (-1)^k \int_0^1 f(u)\, p_k\left((1+\cos(\pi(1-u)))/2 | w_{\mathcal{C}[0,1]}\right) \mathrm{d}u \end{aligned}$$

$$= \begin{cases} \int_0^1 f(u)\,\mathrm{d}u & \text{for} \quad k = 0, \\ \int_0^1 f(u)\sqrt{2}\cos(k\pi u)\,\mathrm{d}u & \text{for} \quad k \in \mathbb{N}. \end{cases}$$

Consequently, we have the following results.

**Theorem 6.1.** *The functions*

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

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

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

*Consequently, by Lemma 5.1, $f$ can be written as*

$$f(u) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2}\cos(k\pi u) \text{ a.e. on } (0,1). \tag{6.3}$$

## 6.2 Fourier series representation

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

$$\begin{aligned} \varphi_0(x) &= 1 \\ \varphi_{2k-1}(x) &= \sqrt{2}\sin(k\pi x),\ \varphi_{2k}(x) = \sqrt{2}\cos(k\pi x),\ k \in \mathbb{N}. \end{aligned} \tag{6.4}$$

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.,

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

It is a classical Fourier analysis result that

**Theorem 6.2**. *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 6.1 this result can be proved somewhat easier, as follows.

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

$$\lim_{n\to\infty} \frac{1}{2}\int_{-1}^{1} (g(x) - g_n(x))^2 \mathrm{d}x = 0, \tag{6.5}$$

where

$$g_n(x) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi x) + \sum_{k=1}^{n} \beta_k \sqrt{2}\sin(k\pi x) \tag{6.6}$$

with Fourier coefficients

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

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

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

and

$$\begin{aligned} f_n(u) &= g_n(2u-1) \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi(2u-1)) + \sum_{k=1}^{n} \beta_k \sqrt{2}\sin(k\pi(2u-1)) \\ &= \alpha_0 + \sum_{k=1}^{n} \alpha_k (-)^k \sqrt{2}\cos(2k\pi u) + \sum_{k=1}^{n} \beta_k (-)^k \sqrt{2}\sin(2k\pi u) \,. \quad (6.7) \end{aligned}$$

Thus, (6.5) is true if and only

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

Theorem 6.2 follows now from the following result, which will be proved in the appendix to this chapter.

**Theorem 6.3**. *The functions* $\overline{\varphi}_0(u) = 1$, $\overline{\varphi}_{2k-1}(u) = \sqrt{2}\sin(2k\pi u)$, $\overline{\varphi}_{2k}(u) = \sqrt{2}\cos(2k\pi u)$, $k \in \mathbb{N}$, *form a complete orthonormal sequence in* $L^2(0,1)$. *Consequently, any function* $f \in L^2(0,1)$ *has the series representation*

$$\begin{aligned} f(u) &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_{2k-1}\sqrt{2}\sin(2k\pi u) \\ &\quad + \sum_{k=1}^{\infty} \gamma_{2k}\sqrt{2}\cos(2k\pi u) \;\; a.e. on \;(0,1), \qquad (6.8) \end{aligned}$$

*where* $\gamma_0 = \int_0^1 f(u)\mathrm{d}u$, $\gamma_{2k-1} = \int_0^1 f(u)\sqrt{2}\sin(2k\pi u)\mathrm{d}u$, $\gamma_{2k} = \int_0^1 f(u)\sqrt{2}\cos(2k\pi u)\mathrm{d}u$, *and* $\sum_{k=0}^{\infty}\gamma_k^2 = \int_0^1 f(u)^2\mathrm{d}u < \infty$.

As indicated before, Theorem 6.2 is a corollary of Theorem 6.3. The other way around is also true, i.e., Theorem 6.3 is a corollary of Theorem 6.2. To see this, let $f \in L^2(0,1)$ be arbitrary, and denote $g(x) = f(x+1)/2)$. Obviously, $g \in L^2(-1,1)$. Then with $g_n(x)$ and $f_n(u)$ as in (6.6) and (6.7), respectively, we have

$$\int_0^1 (f(u) - f_n(u))^2 \,\mathrm{d}u = \frac{1}{2}\int_{-1}^1 (g(x) - g_n(x))^2 \,\mathrm{d}x \to 0$$

as $n \to \infty$. Therefore, in order to make this book self-contained, I need to provide either the proof of Theorem 6.2 or the proof of Theorem 6.3.

Moreover, Theorem 6.1 is also a corollary of Theorem 6.2. To see this, let $f(u) \in L^2(0,1)$ be arbitrary, and denote $g(x) = f(|x|)$. Then $g(x) \in L^2(-1,1)$, with Fourier coefficients

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

Hence, it follows from Theorem 6.2 that

$$\begin{aligned}
&\lim_{n\to\infty}\int_0^1 \left(f(u) - \alpha_0 - \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi u)\right)^2 \mathrm{d}u \qquad (6.9)\\
&= \frac{1}{2}\lim_{n\to\infty}\int_{-1}^{1} \left(f(|x|) - \alpha_0 - \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi x)\right)^2 \mathrm{d}x \\
&= 0.
\end{aligned}$$

Is follows now from (6.9) and Lemma 5.1 that $f(u) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2}\cos(k\pi u)$ a.e. on $(0,1)$, where $\alpha_0 = \int_0^1 f(u)\mathrm{d}u$ and $\alpha_k = \int_0^1 \sqrt{2}\cos(k\pi u)\, f(u)\mathrm{d}u$ for $k \in \mathbb{N}$, which is just the result in Theorem 6.1.

## 6.3 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

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

Hence by Theorem 6.2, $\lim_{n\to\infty}\frac{1}{2}\int_{-1}^{1}\left(f(x)-\sum_{k=1}^{n}\beta_k\sqrt{2}\sin(k\pi x)\right)^2\mathrm{d}x=0$, which by the condition $f(x)=-f(-x)$ implies

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

where

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

Moreover, note that

$$\begin{aligned}&\int_0^1\sqrt{2}\sin(k\pi u)\sqrt{2}\sin(m\pi u)\,\mathrm{d}u\\&=\int_0^1\cos((k-m)\pi u)\,\mathrm{d}u-\int_0^1\cos((k+m)\pi u)\,\mathrm{d}u\\&=I(k=m),\end{aligned}$$

so that the sequence $\{\sqrt{2}\sin(k\pi u)\}_{k=1}^{\infty}$ is orthonormal. Thus, we have the following corollary of Theorem 6.2.

**Theorem 6.4**. *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*

$$f(u)=\sum_{k=1}^{\infty}\beta_k\sqrt{2}\sin(k\pi u)\quad a.e.\ on\ (0,1), \tag{6.10}$$

*where* $\beta_k=\int_0^1 f(u)\sqrt{2}\sin(k\pi u)\mathrm{d}u$ with $\sum_{k=1}^{\infty}\beta_k^2=\int_0^1 f(u)^2\mathrm{d}u<\infty$.

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.

## 6.4 How well does the cosine series fit?

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 exactly, as

$$\begin{aligned}\alpha_0 &= \int_0^1 f(u)\mathrm{d}u = 1, \\ \alpha_k &= \int_0^1 f(u)\sqrt{2}\cos(k\pi u)\mathrm{d}u \\ &= \begin{cases} -6\sqrt{2}(k\pi)^{-2} & \text{if } k \geq 2 \text{ is even,} \\ -2\sqrt{2}(k\pi)^{-2} & \text{if } k \geq 1 \text{ is odd.} \end{cases} \end{aligned} \tag{6.11}$$

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 approximations $f_n(u) = 1 + \sum_{k=1}^n \alpha_k\sqrt{2}\cos(k\pi u)$ (dotted curve) for $n = 10$ and $n = 20$.

*Insert*
Figure 6.1: $f(u) = u(4-3u)$ compared with $f_{10}(u)$
*about here.*

*Insert*
Figure 6.2: $f(u) = u(4-3u)$ compared with $f_{20}(u)$
*about here.*

We see that $f_n(u)$ approximates $f(u)$ quite well, except in the tails of $f_n(u)$. 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 $n$.

Note that in this case

$$\sup_{0\leq u\leq 1}|f(u) - f_n(u)| \leq \sqrt{2}\sum_{k=n+1}^{\infty}|\alpha_k| < \frac{12}{\pi^2}\sum_{k=n+1}^{\infty}k^{-2} \to 0$$

as $n \to \infty$, where the second inequality follows from (6.11). This implies that in the cosine series representation

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

both sides of this equation are uniformly continuous on $[0,1]$, hence we may drop the caveat "a.e." and conclude that

$$f(u) \equiv 1 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos(k\pi u) \text{ on } [0,1].$$

Consequently, for $u \in (0,1)$ and $\varepsilon \neq 0$ so small that $u + \varepsilon \in (0,1)$ we have

$$\begin{aligned} f'(u) &\equiv \lim_{\varepsilon\to 0} \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \frac{\cos(k\pi(u+\varepsilon)) - \cos(k\pi u)}{\varepsilon} \\ &= f_n'(u) + \lim_{\varepsilon\to 0} \sum_{k=n+1}^{\infty} \alpha_k \sqrt{2} \frac{\cos(k\pi(u+\varepsilon)) - \cos(k\pi u)}{\varepsilon} \\ &= f_n'(u) - \pi \lim_{\varepsilon\to 0} \sum_{k=n+1}^{\infty} k\alpha_k \frac{1-\cos(k\pi\varepsilon)}{k\pi\varepsilon} \sqrt{2}\cos(k\pi u) \\ &\quad -\pi \lim_{\varepsilon\to 0} \sum_{m=n+1}^{\infty} m\alpha_m \frac{\sin(m\pi\varepsilon)}{m\pi\varepsilon} \sqrt{2}\sin(m\pi u) \end{aligned}$$

However, it follows from (6.11) that

$$\sum_{m=n+1}^{\infty} m|\alpha_m| \geq \left(2\sqrt{2}/\pi^2\right) \sum_{m=n+1}^{\infty} m^{-1} = \infty.$$

Consequently, we are not allowed to bring the limit operation $\lim_{\varepsilon\to 0}$ inside the summations, because the conditions of the dominated convergence theorem for doing this are not met. On the other hand, it follows from (6.11) that $\sum_{m=1}^{\infty}(m\alpha_m)^2 < \infty$, so that by Theorems 6.1 and 6.4 and the dominated convergence theorem,

$$\begin{aligned} \int_0^1 \left(f'(u) - f_n'(u)\right)^2 \mathrm{d}u &\leq 2\pi^2 \lim_{\varepsilon\to 0} \sum_{k=n+1}^{\infty} (k\alpha_k)^2 \left(\frac{1-\cos(k\pi\varepsilon)}{k\pi\varepsilon}\right)^2 \\ &\quad +2\pi^2 \lim_{\varepsilon\to 0} \sum_{k=n+1}^{\infty} (k\alpha_k)^2 \left(\frac{\sin(m\pi\varepsilon)}{m\pi\varepsilon}\right)^2 \\ &= 2\pi^2 \sum_{k=n+1}^{\infty} (k\alpha_k)^2 \left(\lim_{\varepsilon\to 0} \frac{1-\cos(k\pi\varepsilon)}{k\pi\varepsilon}\right)^2 \end{aligned}$$

$$+2\pi^2 \sum_{k=n+1}^{\infty} (k\alpha_k)^2 \left( \lim_{\varepsilon \to 0} \frac{\sin(m\pi\varepsilon)}{m\pi\varepsilon} \right)^2$$
$$= 2\pi^2 \sum_{k=n+1}^{\infty} (k\alpha_k)^2 \to 0 \text{ as } n \to \infty,$$

hence,

$$f'(u) = -\sum_{k=1}^{\infty} \alpha_k k\pi\sqrt{2}\sin(k\pi u) \text{ a.e. on } (0,1). \tag{6.12}$$

Note that $-k\pi\alpha_k = -k\pi\int_0^1 f(u)\sqrt{2}\cos(k\pi u)\mathrm{d}u = \int_0^1 f'(u)\sqrt{2}\sin(k\pi u)\mathrm{d}u$, so that (6.12) is just the result of Theorem 6.4.

## 6.5 How well does the Fourier series fit?

To check how well the Fourier series (c.f. Theorem 6.3) fit, consider the same function $f(u) = u(4-3u)$ on $[0,1]$ as before. In this case the exact Fourier coefficients involved are

$$\begin{aligned}
\gamma_0 &= \int_0^1 f(u)\mathrm{d}u = 1, \\
\gamma_{2k} &= \int_0^1 f(u)\sqrt{2}\cos(2k\pi u)\mathrm{d}u = -\left(3/\sqrt{2}\right)(k\pi)^{-2}, \\
\gamma_{2k-1} &= \int_0^1 f(u)\sqrt{2}\sin(2k\pi u)\mathrm{d}u = -\left(\sqrt{2}k\pi\right)^{-1}.
\end{aligned}$$

In the following figures the function $f(u) = u(4-3u)$ is compared with its Fourier series approximations

$$f_n(u) = 1 + \sum_{k=1}^{n/2} \gamma_{2k}\sqrt{2}\cos(2k\pi u) + \sum_{k=1}^{n/2} \gamma_{2k-1}\sqrt{2}\sin(2k\pi u)$$

(dotted curve) for $n = 10$ and $n = 20$.

*Insert*
Figure 6.3: $f(u) = u(4-3u)$ compared with $f_{10}(u)$
*about here.*

*Insert*
Figure 6.4: $f(u) = u(4-3u)$ compared with $f_{20}(u)$
*about here.*

Clearly, the fit of the Fourier series approximations is pretty bad compared with the cosine series approximations, even in the tails. This bad fit may be due to the slower rate of convergence to zero of $\gamma_{2k-1}$, i.e., $\gamma_{2k-1} = O(k^{-1})$, compared with $\alpha_{2k-1} = O(k^{-2})$ in the cosine case, whereas $\gamma_{2k}$ and $\alpha_{2k}$ are both of order $O(k^{-2})$. Consequently, in the present case we have $\int_0^1 (f(u) - f_n(u))^2 \mathrm{d}u = O(n^{-2})$, as is not hard to verify, whereas in the cosine case $\int_0^1 (f(u) - f_n(u))^2 \mathrm{d}u = O(n^{-4})$.

## 6.6 How well, or bad, does the sine series fit?

Recall that the derivative $f'(u) = 4 - 6u$ of $f(u) = u(4-3u)$ has series representation (6.12), and that with

$$f'_n(u) = -\sum_{k=1}^{n} \alpha_k k\pi\sqrt{2}\sin(k\pi u), \tag{6.13}$$

where the $\alpha_k$'s are given in (6.11), we have, by Theorem 6.4, that

$$\lim_{n\to\infty} \int_0^1 (f'(u) - f'_n(u))^2 \mathrm{d}u = 0.$$

Also recall that the representation (6.12) and its approximation (6.13) suffer from tail problems, i.e., $f'_n(0) = f'_n(1) = 0$ whereas $f'(0) = 4$, $f'(1) = -2$.

In the following two figures the derivative $f'(u) = 4 - 6u$ and its approximation $f'_n(u)$ are compared, for $n = 10$ and $n = 20$.

*Insert*
Figure 6.5: $f'(u) = 4 - 6u$ compared with $f'_{10}(u)$
*about here.*

*Insert*
Figure 6.6: $f'(u) = 4 - 6u$ compared with $f'_{20}(u)$
*about here.*

The fit in the mid section of $(0,1)$ in the case $n = 10$ is not great, but it improves somewhat for $n = 20$, as Theorem 6.4 predicts. To see how much the fit improves for larger $n$, consider the case $n = 50$, in Figure 6.7.

*Insert*
Figure 6.7: $f'(u) = 4 - 6u$ compared with $f'_{50}(u)$
*about here.*

What we see happening in these pictures is that $\lim_{n\to\infty} f_n^{'}(u) = f'(u)$ a.e. on $(0,1)$, as predicted by Theorem 6.4, but

$$\lim_{n\to\infty} \sup_{0\leq u\leq 1} |f_n^{'}(u) - f'(u)| > 0.$$

An explanation for the tail behavior of the sine series representation is the following.

**Lemma 6.1**. *For $k \in \mathbb{N}$ and $u \in [0,1]$,*

$$\sin(k\pi u) = \sin(\pi u).P_{k-1}^{s}(\cos(\pi u)),$$

*where*

$$P_{k-1}^{s}(z) = k.z^{k-1} + \sum_{m=2}^{[(k+1)/2]} \binom{k}{2m-1} (-1)^{m-1} z^{k-2m+1} \left(1 - z^2\right)^{m-1}$$

*is a polynomial of order $k-1$ in $z \in [-1,1]$, with $[x]$ denoting the largest integer $\leq x$. Moreover,*

$$\cos(k\pi u) = P_k^c(\cos(\pi u)),$$

*where*

$$P_k^c(z) = z^k + \sum_{m=1}^{[k/2]} \binom{k}{2m} (-1)^m z^{k-2m} \left(1 - z^2\right)^m$$

*is a polynomial of order $k$ in $z \in [-1,1]$.*

Therefore, the sine series representation (6.10) in Theorem 6.4 reads

$$f(u) = \sqrt{2}\sin(\pi u) \sum_{k=1}^{\infty} \beta_k P_{k-1}^{s}(\cos(\pi u)) \text{ a.e. on } (0,1). \qquad (6.14)$$

Clearly, the closer the distance of $u$ to 0 or 1, the more dominant the factor $\sqrt{2}\sin(\pi u)$ in the sine series representation (6.14) becomes, which is one of the explanations of the tail behavior in Figure 6.7.

Another explanation is the following. Note that

$$\begin{aligned}\lim_{u\downarrow 0}\frac{\sin(k\pi u)}{\sin(\pi u)} &= P^s_{k-1}(1) = k,\\ \lim_{u\uparrow 1}\frac{\sin(k\pi u)}{\sin(\pi u)} &= P^s_{k-1}(-1) = k(-1)^{k-1},\end{aligned}$$

hence it follows from (6.13), with $\alpha_k$ given by (6.11), that

$$\begin{aligned}\lim_{u\downarrow 0}\frac{f_n'(u)}{n\sqrt{2}\sin(\pi u)} &= -\frac{\pi}{n}\sum_{k=1}^{n}\alpha_k k^2\\ &= -\frac{\pi}{n}\sum_{k=1}^{[n/2]}\alpha_{2k}(2k)^2 - \frac{\pi}{n}\sum_{k=1}^{[(n+1)/2]}\alpha_{2k-1}(2k-1)^2\\ &= \pi^{-1}2\sqrt{2}\left(\frac{3[n/2]+[(n+1)/2]}{n}\right)\\ &\to \pi^{-1}4\sqrt{2} \text{ as } n\to\infty,\end{aligned}$$

so that for large $n$ and $u$ close to 0,

$$f_n'(u) \approx n.8\pi^{-1}\sin(\pi u) \tag{6.15}$$

Similarly,

$$\lim_{u\uparrow 1}\frac{f_n'(u)}{n\sqrt{2}\sin(\pi u)} \to -\pi^{-1}4\sqrt{2} \text{ as } n\to\infty,$$

so that for large $n$ and $u$ close to 1,

$$f_n'(u) \approx -n.8\pi^{-1}\sin(\pi u). \tag{6.16}$$

The approximations (6.15) and (6.16) suggest that the rate of convergence of $f_n'(u)$ to $f'(u)$ slows down the closer $u$ gets to 0 or 1, which is what happens in Figure 6.7.

# 6.7 Proofs

## 6.7.1 Theorem 6.3

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 $\int_0^1 f(u)\mathrm{d}u = 0$, 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 $F(u) = \int_0^u f(x)\mathrm{d}x$ with $f(u) = F'(u)$. It will be shown that $C_{0,1}[0,1] \subset \overline{\text{span}}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

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

*Step 3.* Finally, it will be shown that every simple function in $L^2(0,1)$ can be written as a limit of a sequence of functions in $C[0,1]$. Since functions are Borel measurable if an only if they are pointwise limits of a sequence of simple functions, it follows that any Borel measurable function is a pointwise limit of a sequence of functions contained in $C[0,1]$. Hence, $L^2(0,1)$ is the closure of $C[0,1]$, so that $L^2(0,1) = \overline{\text{span}}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

**Proof of Step 1**

Let $f_n(u)$ and $f(u)$ be the same as in Theorem 6.1, except that due to the condition $\int_0^1 f(u)\mathrm{d}u = 0$, $\alpha_0 = 0$, and let $F_n(u) = \int_0^u f_n(x)\mathrm{d}x$. 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

$$\begin{aligned} \sup_{0\le u\le 1} |F(u) - F_n(u)| &\le \int_0^1 |f(x) - f_n(x)|\,\mathrm{d}x \\ &\le \sqrt{\int_0^1 (f(x) - f_n(x))^2\,\mathrm{d}x} = o(1). \quad (6.17) \end{aligned}$$

Next, observe that

$$\int_0^1 \sqrt{2}\sin\left(\left(2k-1\right)\pi u\right)\mathrm{d}u = \frac{-2\sqrt{2}}{\left(2k-1\right)\pi} = \gamma_{0,k}$$
$$\int_0^1 \sqrt{2}\sin\left(\left(2k-1\right)\pi u\right)\sqrt{2}\cos\left(\left(2m-1\right)\pi u\right)\mathrm{d}u = 0$$
$$\begin{aligned}\int_0^1 &\sqrt{2}\sin\left(\left(2k-1\right)\pi u\right)\sqrt{2}\cos\left(2m\pi u\right)\mathrm{d}u \\ &= \frac{-2}{\left(2\left(k+m\right)-1\right)\pi} + \frac{-2}{\left(2\left(k-m\right)-1\right)\pi} \\ &= -\frac{2}{\pi}\frac{4k-2}{\left(2\left(k+m\right)-1\right)\left(2\left(k-m\right)-1\right)} \\ &= -\frac{2}{\pi}\frac{k-1/2}{\left(k-1/2\right)^2-m^2} = \gamma_{m,k}\end{aligned}$$

Hence

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

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

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

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

$$\int_0^1\left(\widetilde{F}_n\left(u\right)-F_n\left(u\right)\right)^2\mathrm{d}u = \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\sum_{k=1}^{[(n+1)/2]}\frac{\alpha_{2k-1}}{m^2-\left(k-1/2\right)^2}\right)^2$$

$$\begin{aligned}
&\leq \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\sum_{k=1}^{[(n+1)/2]}\frac{|\alpha_{2k-1}|}{m^2-(k-1/2)^2}\right)^2 \\
&\leq \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\frac{\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{m^2-([n/2])^2}\right)^2 \\
&= \frac{1}{\pi^4}\sum_{m=N+1}^{\infty}\left(\frac{\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{(m-[n/2])(m+[n/2])}\right)^2 \\
&\leq \frac{1}{4\pi^4}\sum_{m=N+1}^{\infty}\left(\frac{\frac{1}{[n/2]}\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{m-[n/2]}\right)^2 \\
&\leq \frac{1}{4\pi^4}\sum_{m=[n/2]+1}^{\infty}\left(\frac{\frac{1}{[n/2]}\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|}{m-[n/2]}\right)^2 \\
&= \frac{1}{4\pi^4}\left(\sum_{m=1}^{\infty}\frac{1}{m^2}\right)\left(\frac{1}{[n/2]}\sum_{k=1}^{[(n+1)/2]}|\alpha_{2k-1}|\right)^2 \\
&\leq \frac{1}{4\pi^4}\left(\sum_{m=1}^{\infty}\frac{1}{m^2}\right)\frac{1}{[n/2]}\sum_{k=1}^{\infty}\alpha_{2k-1}^2 \\
&= O(1/n). \qquad (6.18)
\end{aligned}$$

Hence by (6.17) and (6.18),

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

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

**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 $F(u) = \int_0^u f(x)\mathrm{d}x$ and

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

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 $u_1,u_2\in[0,1]$, and $\overline{B}\backslash B\neq\emptyset$,

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

hence

$$\left|\inf_{x\in\overline{B}}|x-u_2|-\inf_{x\in\overline{B}}|x-u_1|\right|\leq|u_2-u_1| \tag{6.19}$$

and similarly,

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

In the case $\overline{B}=B$, only (6.19) applies.

Suppose again that $\overline{B}\backslash B \neq \emptyset$. 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 $u \in \overline{B}\backslash B$, $\inf_{x\in\overline{B}}|x-u| = 0$ and $\inf_{x\in\overline{B}\backslash B}|x-u| = 0$, hence $f_n(u) = 0$. In the case $\overline{B} = B$ we may skip the latter step. Moreover, 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

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

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 $B$ in $[0,1]$, $I(u \in B) \in \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$ 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, so that such a sequence is a Cauchy sequence in $L^2(0,1)$, it follows that $L^2(0,1) = \text{span}(\{\overline{\varphi}_n\}_{n=0}^{\infty})$.

Finally, (6.8) follows from Lemma 5.1

### 6.7.2 Lemma 6.1

It follows from the well-known sine-cosine formulas that

$$\begin{aligned}\begin{pmatrix}\sin(k\pi u)\\ \cos(k\pi u)\end{pmatrix} &= \begin{pmatrix}\cos(\pi u) & \sin(\pi u)\\ -\sin(\pi u) & \cos(\pi u)\end{pmatrix}\begin{pmatrix}\sin((k-1)\pi u)\\ \cos((k-1)\pi u)\end{pmatrix}\\ &= \begin{pmatrix}\cos(\pi u) & \sin(\pi u)\\ -\sin(\pi u) & \cos(\pi u)\end{pmatrix}^k\begin{pmatrix}0\\ 1\end{pmatrix}.\end{aligned}$$

Using the binomial expansion we can write

$$\begin{aligned}&\begin{pmatrix}\sin(k\pi u)\\ \cos(k\pi u)\end{pmatrix}\\ &= \left(\sin(\pi u)\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} + \cos(\pi u)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)^k\begin{pmatrix}0\\ 1\end{pmatrix}\\ &= \sum_{m=0}^{k}\binom{k}{m}\cos^{k-m}(\pi u)\sin^m(\pi u)\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}^m\begin{pmatrix}0\\ 1\end{pmatrix},\end{aligned}$$

where we should interpret

$$\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}^0 = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}.$$

Moreover, it is easy to verify, by induction, that for $m \in \mathbb{N}$,

$$\begin{aligned}
\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{2m} \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= (-1)^m \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \\
\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{2m-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} &= -(-1)^m \begin{pmatrix} 1 \\ 0 \end{pmatrix}.
\end{aligned}$$

Hence,

$$\begin{aligned}
\cos(k\pi u) &= \sum_{m=0}^{[k/2]} \binom{k}{2m} (-1)^m \cos^{k-2m}(\pi u) \sin^{2m}(\pi u) \\
&= \sum_{m=0}^{[k/2]} \binom{k}{2m} (-1)^m \cos^{k-2m}(\pi u) \left(1 - \cos^2(\pi u)\right)^m \\
&= \cos^k(\pi u) \\
&\quad + \sum_{m=1}^{[k/2]} \binom{k}{2m} (-1)^m \cos^{k-2m}(\pi u) \left(1 - \cos^2(\pi u)\right)^m \qquad (6.20)
\end{aligned}$$

and

$$\begin{aligned}
\sin(k\pi u) &= -\sum_{m=1}^{[(k+1)/2]} \binom{k}{2m-1} (-1)^m \cos^{k-2m+1}(\pi u) \\
&\quad \times \sin^{2m-1}(\pi u) \\
&= \sin(\pi u) \\
&\quad \times \sum_{m=1}^{[(k+1)/2]} \binom{k}{2m-1} (-1)^{m-1} \cos^{k-2m+1}(\pi u) \\
&\quad \times \sin^{2(m-1)}(\pi u) \\
&= \sin(\pi u) \\
&\quad \times \sum_{m=1}^{[(k+1)/2]} \binom{k}{2m-1} (-1)^{m-1} \cos^{k-2m+1}(\pi u) \\
&\quad \times \left(1 - \cos^2(\pi u)\right)^{m-1} \\
&= \sin(\pi u) \left\{ k \cos^{k-1}(\pi u) \right.
\end{aligned}$$

$$+\binom{k}{2m-1}(-1)^{m-1}\cos^{k-2m+1}(\pi u)$$
$$\times\left(1-\cos^2(\pi u)\right)^{m-1}\Bigg\}. \tag{6.21}$$

The theorem under review now follows straightforwardly from (6.20) and (6.21).

# Chapter 7

# Density and distribution functions

## 7.1 General univariate SNP density functions

Let $w(x)$ be a density function with support $(a,b)$, where possibly $a=-\infty$ and/or $b=\infty$. Then for any density $f(x)$ on $(a,b)$,

$$g(x)=\sqrt{f(x)}/\sqrt{w(x)}\in L^2(w), \tag{7.1}$$

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

$$f(x)=w(x)\left(\sum_{m=0}^{\infty}\gamma_m\rho_m(x)\right)^2 \text{ a.e. on } (a,b), \text{ with } \sum_{m=0}^{\infty}\gamma_m^2=\int_a^b f(x)\mathrm{d}x=1. \tag{7.2}$$

The reason for the square in (7.2) is to guarantee that $f(x)$ is non-negative. The "a.e." part in (7.2) follows from Lemma 5.1.

A problem with the series representation (7.2) is that in general the parameters involved are not unique. To see this, note that if we replace the function $g(x)$ in (7.1) by $g_B(x)=(I(x\in B)-I(x\in(a,b)\backslash B))\sqrt{f(x)}/\sqrt{w(x)}$, where $B$ is an arbitrary Borel set, then $g_B(x)\in L^2(w)$ and $\int_a^b g_B(x)^2 w(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x = 1$, so that (7.2) also holds for the sequence

$$\gamma_m = \int_a^b \rho_m(x)g_B(x)w(x)\mathrm{d}x$$

$$= \int_{(a,b)\cap B} \rho_m(x)\sqrt{f(x)}\sqrt{w(x)}\mathrm{d}x - \int_{(a,b)\setminus B} \rho_m(x)\sqrt{f(x)}\sqrt{w(x)}\mathrm{d}x$$

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

$$\gamma_0 = \int_{(a,b)\cap B} \sqrt{f(x)}\sqrt{w(x)}\mathrm{d}x - \int_{(a,b)\setminus B} \sqrt{f(x)}\sqrt{w(x)}\mathrm{d}x,$$

so that the sequence $\gamma_m$ in (7.2) is unique if $\gamma_0$ is maximal. In any case we may without loss of generality assume that $\gamma_0 \in (0,1)$, so that without loss of generality the $\gamma_m$'s can be reparameterized as

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

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

On the other hand, under certain conditions on $f(x)$ the $\delta_m$'s are unique, as will be shown in section 7.4 below.

Summarizing, the following results have been shown.

**Theorem 7.1.** *Let $w(x)$ be a univariate density function with support $(a,b)$, where possibly $a = -\infty$ and/or $b = \infty$, and let $\{\rho_m\}_{m=0}^{\infty}$ be a complete orthonormal sequence in $L^2(w)$, with $\rho_0(x) \equiv 1$. Then for any density $f(x)$ on $(a,b)$ there exist possibly uncountable many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty}\delta_m^2 < \infty$ such that*

$$f(x) = \frac{w(x)\left(1+\sum_{m=1}^{\infty}\delta_m\rho_m(x)\right)^2}{1+\sum_{m=1}^{\infty}\delta_m^2} \text{ a.e. on } (a,b). \tag{7.3}$$

*Moreover, denoting for $n \in \mathbb{N}$,*

$$f_n(x) = \frac{w(x)\left(1+\sum_{m=1}^{n}\delta_m\rho_m(x)\right)^2}{1+\sum_{m=1}^{n}\delta_m^2} \tag{7.4}$$

*we have that $\lim_{n\to\infty} f_n(x) = f(x)$ a.e. on $(a,b)$. Furthermore, it is not hard to verify that*

$$\int_a^b |f(x) - f_n(x)|\mathrm{d}x \le 4\sqrt{\sum_{m=n+1}^{\infty}\delta_m^2} + 2\sum_{m=n+1}^{\infty}\delta_m^2 = o(1), \tag{7.5}$$

*so that with $F(x)$ the c.d.f. of $f(x)$ and $F_n(x)$ the c.d.f. of $f_n(x)$,*

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

Following Gallant and Nychka (1987), I will refer to truncated densities of the type (7.4) as *SNP densities*.

**Remarks:**

1. The rate of convergence to zero of the tail sum $\sum_{m=n+1}^{\infty}\delta_m^2$ depends on the smoothness, or the lack thereof, of the density $f(x)$. Therefore, the question how to choose the truncation order $n$ given an a priori chosen approximation error cannot be answered in general.

2. In the case that the $\rho_m(x)$'s are orthonormal polynomials, the SNP density $f_n(x)$ has to be computed recursively via the corresponding TTRR (5.12), except in the case of Chebyshev polynomials, but that is not too much of a computational burden. However, the computation of the corresponding SNP distribution function $F_n(x)$ is more complicated. See for example Stewart (2004) for SNP distribution functions on $\mathbb{R}$ based on Hermite polynomials, and Bierens (2008) for SNP distribution functions on $[0,1]$ based on Legendre polynomials. Both cases require to recover the coefficients $\ell_{m,k}$ of the polynomials $\overline{p}_k(x|w) = \sum_{m=0}^{k}\ell_{m,k}x^m$, which can be done using the TTRR involved. Then with $P_n(x|w) = (1, \overline{p}_1(x|w), ..., \overline{p}_n(x|w))'$, $Q_n(x) = (1, x, , ..., x^n)'$, $\delta = (1, \delta_1, ..., \delta_n)$, and $L_n$ the lower-triangular matrix consisting of the coefficients $\ell_{m,k}$, we can write

$$f_n(x) = (\delta'\delta)^{-1}w(x)(\delta' P_n(x|w))^2 = (\delta'\delta)^{-1}\delta' L_n Q_n(x)Q_n(x)'w(x)L_n'\delta,$$

hence

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

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

## 7.2 Bivariate SNP density functions

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

$$\int_{a_1}^{b_1} \int_{a_2}^{b_2} g(x, y)^2 w_1(x) w_2(y) \mathrm{d}x \mathrm{d}y < \infty. \tag{7.6}$$

The latter implies that $g_2(y) = \int_{a_1}^{b_1} g(x, y)^2 w_1(x) \mathrm{d}x < \infty$ a.e. on $(a_2, b_2)$, so that for each $y \in (a_2, b_2)$ for which $g_2(y) < \infty$, $g(x, y) \in L^2(w_1)$. Then $g(x, y) = \sum_{m=0}^{\infty} \gamma_m(y) \rho_{1,m}(x)$ a.e. on $(a_1, b_1)$, where $\gamma_m(y) = \int_{a_1}^{b_1} g(x, y) \rho_{1,m}(x) w_1(x) \mathrm{d}x$ with $\sum_{m=0}^{\infty} \gamma_m(y)^2 = \int_{a_1}^{b_1} g(x, y)^2 w_1(x) \mathrm{d}x = g_2(y)$. Because by (7.6), $\int_{a_2}^{b_2} g_2(y) w_2(y) \mathrm{d}y < \infty$, it follows now that for each $y \in (a_2, b_2)$ for which $g_2(y) < \infty$ and all integers $m \geq 0$, $\gamma_m(y) \in L^2(w_2)$, so that $\gamma_m(y) = \sum_{k=0}^{\infty} \gamma_{m,k} \rho_{2,k}(y)$ a.e. on $(a_2, b_2)$, where $\gamma_{m,k} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} g(x, y) \rho_{1,m}(x) \rho_{2,k}(y) w_1(x) w_2(y) \mathrm{d}x \mathrm{d}y$ with $\sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \gamma_{m,k}^2 < \infty$. Hence,

$$g(x, y) = \sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \gamma_{m,k} \rho_{1,m}(x) \rho_{2,k}(y) \text{ a.e. on } (a_1, b_1) \times (a_2, b_2). \tag{7.7}$$

Therefore, it follows similar to Theorem 7.1 that

**Theorem 7.2.** *Given a pair of density functions $w_1(x)$ and $w_2(y)$ with supports $(a_1, b_1)$ and $(a_2, b_2)$, respectively, and given complete orthonormal sequences $\{\rho_{1,m}\}_{m=0}^{\infty}$ and $\{\rho_{2,m}\}_{m=0}^{\infty}$ in $L^2(w_1)$ and $L^2(w_2)$, respectively, with $\rho_{1,0}(x) = \rho_{2,0}(y) \equiv 1$, for every bivariate density $f(x, y)$ on $(a_1, b_1) \times (a_2, b_2)$ there exist possibly uncountable many double arrays $\delta_{m,k}$ satisfying $\sum_{m=0}^{\infty} \sum_{k=0}^{\infty} \delta_{m,k}^2 < \infty$, with $\delta_{0,0} = 1$ by normalization, such that a.e. on $(a_1, b_1) \times (a_2, b_2)$,*

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

For example, let $w_1(x)$ and $w_2(y)$ be standard normal densities and let $\rho_{1,m}(x)$ and $\rho_{2,k}(y)$ Hermite polynomials, i.e., $\rho_{1,k}(x) = \rho_{2,k}(x) = \overline{p}_k(x | w_{\mathcal{N}[0,1]})$.

Then for any density function $f(x,y)$ on $\mathbb{R}^2$ there exists a double array $\delta_{m,k}$ and associated sequence of SNP densities

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

such that $\lim_{n\to\infty} f_n(x,y) = f(x,y)$ a.e. on $\mathbb{R}^2$. This result is used by Gallant and Nychka (1987) to approximate the bivariate error density of the latent variable equations in Heckman's (1979) sample selection model.

Similarly, with $w_1$ and $w_2$ the uniform densities on $[0,1]$ and $\rho_{1,k}(u) = \rho_{2,k}(u)$ the cosine series, for any density $f(u,v)$ on $[0,1]\times[0,1]$ there exists a double array $\delta_{m,k}$ and associated sequence of SNP densities

$$f_n(u,v) = \frac{\left(2\sum_{m=0}^n\sum_{k=0}^n\delta_{m,k}\cos(m\pi u)\cos(k\pi v)\right)^2}{\sum_{k=0}^n\sum_{m=0}^n\delta_{k,m}^2}$$

such that $\lim_{n\to\infty} f_n(u,v) = f(u,v)$ a.e. on $[0,1]\times[0,1]$.

## 7.3 SNP densities and distribution functions on $[0,1]$

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

However, there is an easy trick to avoid this computational burden, by mapping one-to-one any absolutely continuous distribution function $F(x)$ on $(a,b)$ with density $f(x)$ to an absolutely continuous distribution function $H(u)$ with density $h(u)$ on the unit interval, as follows.

Given an absolutely continuous distribution function $G(x)$ with support $\Xi \subseteq \mathbb{R}$, any distribution function $F(x)$ with support $\Xi$ can be written as $F(x) = H(G(x))$, where $H(u) = F(G^{-1}(u))$ is a distribution function on $[0,1]$, with $G^{-1}(u)$ the inverse of $G$. Moreover, if $F$ and $G$ are both absolutely continuous with densities $f$ and $g$, respectively, then $H$ is absolutely continuous with density

$$h(u) = f(G^{-1}(u))\frac{\mathrm{d}G^{-1}(u)}{\mathrm{d}u} = \frac{f(G^{-1}(u))}{g(G^{-1}(u))}, \tag{7.8}$$

and $f(x) = h(G(x))g(x)$. Therefore, $f(x)$ and $F(x)$ can be estimated semi-nonparametrically by estimating $h(u)$ and $H(u)$ semi-nonparametrically.

For example, let $(a,b) = \mathbb{R}$ and choose for $G(x)$ the logistic distribution function $G(x) = 1/(1+\exp(-x))$. Then $g(x) = G(x)(1-G(x))$ and $G^{-1}(u) = \ln(u/(1-u))$, hence

$$h(u) = f(\ln(u/(1-u)))/(u(1-u)).$$

Similarly, if $(a,b) = (0,\infty)$ and if we choose $G(x) = 1-\exp(-x)$, then $g(x) = \exp(-x)$ and $G^{-1}(u) = \ln(1/(1-u))$, so that

$$h(u) = f(\ln(1/(1-u)))/(1-u).$$

The reason for this transformation rather than using the results of Theorem 7.1 directly is that there exist closed form expressions for SNP densities and their distribution functions on the unit interval. For example, it follows from (5.22)-(5.23), with the latter replaced by

$$\begin{aligned} \rho_m(u) &= (-1)^m\sqrt{2}\cos(m.\arccos(2u-1)) \\ &= \sqrt{2}\cos\left(m\pi\left(1-\pi^{-1}\arccos(2u-1)\right)\right) \end{aligned} \tag{7.9}$$

for $m \in \mathbb{N}$, and Theorem 7.1 that

**Theorem 7.3.** *For every density $h(u)$ on $[0,1]$ with corresponding c.d.f. $H(u)$ there exist possibly uncountable many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty}\delta_m^2 < \infty$ such that $h(u) = \lim_{n\to\infty} h_n(u)$ a.e. on $[0,1]$, where*

$$h_n(u) = \frac{\left(1+\sum_{m=1}^{n}(-1)^m\delta_m\sqrt{2}\cos\left(m.\arccos(2u-1)\right)\right)^2}{\pi\sqrt{u\left(1-u\right)}\left(1+\sum_{m=1}^{n}\delta_m^2\right)}, \tag{7.10}$$

*and*

$$\lim_{n\to\infty}\sup_{0\le u\le 1}\left|H_n\left(1-\pi^{-1}\arccos(2u-1)\right)-H(u)\right| = 0,$$

*where*

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

$$+ 2\sum_{k=2}^{n}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin\left((k+m)\pi u\right)}{(k+m)\pi}$$
$$\left. + 2\sum_{k=2}^{n}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin\left((k-m)\pi u\right)}{(k-m)\pi}\right]. \tag{7.11}$$

Moreover, with $w(x)$ the uniform density on $[0, 1]$ and $\rho_m(x)$ the cosine sequence it follows from Theorem 7.1 that

**Theorem 7.4.** *For every density $h(u)$ on $[0, 1]$ with corresponding c.d.f. $H(u)$ there exist possibly uncountable many sequences $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty}\delta_m^2 < \infty$ such that $h(u) = \lim_{n\to\infty} h_n(u)$ a.e. on $[0, 1]$, where*

$$h_n(u) = \frac{\left(1 + \sum_{m=1}^{n}\delta_m\sqrt{2}\cos\left(m\pi u\right)\right)^2}{1 + \sum_{m=1}^{n}\delta_m^2}, \tag{7.12}$$

*and $\lim_{n\to\infty}\sup_{0\le u\le 1}|H_n(u) - H(u)| = 0$, where $H_n(u)$ is defined by (7.11).*

The claim that $H_n(u)$ is defined by (7.11) follows straightforwardly from (7.12) and the well-known equality $\cos(a)\cos(b) = (\cos(a+b)+\cos(a-b))/2$, and the same applies to the result for $H(u)$ in Theorem 7.3.

Note that the results in Theorems 7.3 and 7.4 hold for the same density $h(u)$, regardless its shape. However, the SNP density $h_n(u)$ in Theorem 7.3 is in general U-shaped with $h_n(0) = h_n(1) = \infty$, whereas for fixed $n$ ,the SNP density $h_n(u)$ in Theorem 7.4 is bounded on $[0, 1]$.

Moreover, the uniform convergence of $H_n(u)$ to $H(u)$ implies that exactly and uniformly on $[0, 1]$,

$$H(u) \equiv u$$
$$+\frac{1}{1 + \sum_{m=1}^{\infty}\delta_m^2}\left[2\sqrt{2}\sum_{k=1}^{\infty}\delta_k\frac{\sin\left(k\pi u\right)}{k\pi} + \sum_{m=1}^{\infty}\delta_m^2\frac{\sin\left(2m\pi u\right)}{2m\pi}\right.$$
$$+ 2\sum_{k=2}^{\infty}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin\left((k+m)\pi u\right)}{(k+m)\pi}$$
$$\left. + 2\sum_{k=2}^{\infty}\sum_{m=1}^{k-1}\delta_k\delta_m\frac{\sin\left((k-m)\pi u\right)}{(k-m)\pi}\right]. \tag{7.13}$$

The tail behavior of $h(u)$ in (7.8) depends on the a priori chosen distribution function $G$ and the properties of the density $f$. In particular, it is possible that

$$\lim_{u\downarrow 0} h(u) = \infty \text{ and/or } \lim_{u\uparrow 1} h(u) = \infty, \tag{7.14}$$

depending on the choice of $G$. Clearly, in the case (7.14), $h_n(u)$ in Theorem 7.4 will be a poor approximation of $h(u)$ for $u$ close to 0 or 1.

Note that the results in Theorem 7.3 correspond to the choice $G_*(x) = \Lambda(G(x))$ for $G$, where $G$ in the right-hand side is the same as before and $\Lambda(u) = 1 - \pi^{-1}\arccos(2u-1)$. C.f. (5.22). Then $H^*(u) = H(\Lambda(u))$ with density

$$h^*(u) = \frac{h(1-\pi^{-1}\arccos(2u-1))}{\pi\sqrt{u(1-u)}}. \tag{7.15}$$

Replacing $h$ in (7.15) by (7.12) then yields the SNP density (7.10) in Theorem 7.3. The latter may do a better job in approximating $h(u)$ in the case (7.14), although it is still possible that now

$$\lim_{u\downarrow 0}\sqrt{u\,(1-u)}h(u) = \infty \text{ and/or } \lim_{u\uparrow 1}\sqrt{u\,(1-u)}h(u) = \infty.$$

The problem of how to choose $G$ in order to control the tail behavior of $h(u)$ will be addressed in the next chapter.

## 7.4 Uniqueness of the series representation

The density $h(u)$ in Theorem 7.4 can be written as $h(u) = \eta(u)^2/\int_0^1 \eta(v)^2 \mathrm{d}v$, where

$$\eta(u) = 1 + \sum_{m=1}^{\infty}\delta_m\sqrt{2}\cos(m\pi u) \text{ a.e. on } (0,1). \tag{7.16}$$

Moreover, recall that in general,

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

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

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

$$\eta(u) = (I(u \in B) - I(u \in [0,1]\backslash B))\sqrt{h(u)}\sqrt{1 + \sum_{m=1}^{\infty} \delta_m^2} \tag{7.17}$$

Similarly, given this Borel set $B$ and the corresponding $\delta_m$'s, the SNP density (7.12) can be written as $h_n(u) = \eta_n(u)^2 / \int_0^1 \eta_n(v)^2 \mathrm{d}v$, where

$$\begin{aligned} \eta_n(u) &= 1 + \sum_{m=1}^{n} \delta_m \sqrt{2}\cos(m\pi u) \\ &= (I(u \in B) - I(u \in [0,1]\backslash B))\sqrt{h_n(u)}\sqrt{1 + \sum_{m=1}^{n} \delta_m^2} \end{aligned} \tag{7.18}$$

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

$$I(u \in B) - I(u \in [0,1]\backslash B) = \frac{\eta_n(u)}{\sqrt{h_n(u)}\sqrt{1 + \sum_{m=1}^{n} \delta_m^2}} \tag{7.19}$$

is continuous on $(u_0 - \varepsilon_n(u_0), u_0 + \varepsilon_n(u_0)) \cap (0,1)$. Substituting (7.19) in (7.17) 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 (7.16), $\int_0^1 \eta(u)\mathrm{d}u = 1$. Thus, $\eta(u) > 0$ on $(0,1)$, so that by (7.17), $I(u \in B) - I(u \in [0,1]\backslash B) = 1$ on $(0,1)$.

Consequently, the following result holds.

**Theorem 7.5**. *For every continuous density $h(u)$ on $(0,1)$ with support $(0,1)$ the sequence $\{\delta_m\}_{m=1}^{\infty}$ in Theorem 7.4 is unique, with*

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

As is easy to verify, the same argument applies to the more general densities considered in Theorem 7.1:

**Theorem 7.6**. *Let the conditions of Theorem 7.1 be satisfied. In addition, let the density $w(x)$ and the orthonormal functions $\rho_m(x)$ be continuous on $(a,b)$.*[1] *Then every continuous and positive density $f(x)$ on $(a,b)$ has a unique series representation* (7.3), *with*

$$\delta_m = \frac{\int_a^b \rho_m(x)\sqrt{w(x)}\sqrt{f(x)}\mathrm{d}x}{\int_a^b \sqrt{w(x)}\sqrt{f(x)}\mathrm{d}x},\ m \in \mathbb{N}.$$

Moreover, note that Theorem 7.3 is a special case of Theorem 7.1, with $[a,b]=[0,1]$, $f(u)=h(u), w(u)=1/(\pi\sqrt{u(1-u)})$ and $\rho_m(u)=(-1)^m\sqrt{2}\cos(m.\arccos(2u-1))$ for $m \in \mathbb{N}$, hence it follows from Theorem 7.6 that the following result holds.

**Theorem 7.7.** *For every continuous and positive density $h(u)$ on $(0,1)$ the $\delta_m$'s in Theorem 7.3 are unique and given by*

$$\delta_m = (-1)^m \frac{\int_0^1 \sqrt{2}\cos(m.\arccos(2u-1))\,(u\,(1-u))^{-1/4}\sqrt{h(u)}\mathrm{d}u}{\int_0^1 (u\,(1-u))^{-1/4}\sqrt{h(u)}\mathrm{d}u},\ m \in \mathbb{N}.$$

Finally, it is left to the reader to verify that if the density $f(x,y)$ in Theorem 7.2 is continuous and positive on $(a_1,b_1)\times(a_2,b_2)$ then the double array $\delta_{m,k}$ involved, with normalization $\delta_{0,0}=1$, is unique.

## 7.5 Uniform continuity of SNP distribution functions

As we have seen in Theorem 7.1, given a univariate density function $w(x)$ with support $(a,b)$, where possibly $a=-\infty$ and/or $b=\infty$, and with $\{\rho_m\}_{m=0}^{\infty}$ a

[1] The latter is the case if we choose $\rho_m(x)=\overline{p}(x|w)$.

complete orthonormal sequence in $L^2(w)$, with $\rho_0(x) \equiv 1$, any density $f(x)$ on $(a,b)$ can be written as

$$f(x|\boldsymbol{\delta}) = \frac{w(x)\left(1+\sum_{m=1}^{\infty}\delta_m\rho_m(x)\right)^2}{1+\sum_{k=1}^{\infty}\delta_k^2} \text{ a.e. on } (a,b),$$

where $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=1}^{\infty}\delta_m^2 < \infty$.

Moreover, it has been shown in Theorem 2.7 that the space

$$\mathbb{R}^{\infty} = \left\{\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty}\delta_m^2 < \infty\right\}$$

endowed with the innerproduct

$$\langle\boldsymbol{\delta}_1,\boldsymbol{\delta}_2\rangle = \sum_{m=1}^{\infty}\delta_{1,m}\delta_{2,m},\ \ \boldsymbol{\delta}_i = \{\delta_{i,m}\}_{m=1}^{\infty},$$

and associated norm $||\boldsymbol{\delta}|| = \sqrt{\langle\boldsymbol{\delta},\boldsymbol{\delta}\rangle}$ and metric $||\boldsymbol{\delta}_1 - \boldsymbol{\delta}_2||$ is a Hilbert space. Then the following result holds.

**Theorem 7.8**. *For any pair $\boldsymbol{\delta}$ and $\boldsymbol{\delta}_*$ in $\mathbb{R}^{\infty}$ such that $||\boldsymbol{\delta}_* - \boldsymbol{\delta}|| < 1$,*

$$\int_a^b |f(x|\boldsymbol{\delta}) - f(x|\boldsymbol{\delta}_*)|\,\mathrm{d}z \le 5||\boldsymbol{\delta}_* - \boldsymbol{\delta}||,$$

*which trivially implies that the corresponding distribution function $F(x|\boldsymbol{\delta}) = \int_a^x f(z|\boldsymbol{\delta})\mathrm{d}z$ is uniformly continuous on $\mathbb{R}^{\infty}$, i.e.*

$$\sup_{a<x<b}|F(x|\boldsymbol{\delta}) - F(x|\boldsymbol{\delta}_*)| \le 5||\boldsymbol{\delta}_* - \boldsymbol{\delta}||.$$

**Proof.** For notational convenience I will assume that $a = 0$, $b = 1$, $w(x) = 1$ and $\rho_m(x)$, $m \in \mathbb{N}$, with $\rho_0(x) \equiv 1$ any complete orthonormal sequence in $L^2(0,1)$.

First, observe that

$$f(x|\boldsymbol{\delta}) - f(x|\boldsymbol{\delta}_*) = \frac{\left(1+\sum_{m=1}^{\infty}\delta_m\rho_m(x)\right)^2}{1+\sum_{m=1}^{\infty}\delta_m^2} - \frac{\left(1+\sum_{m=1}^{\infty}\delta_{*,m}\rho_m(x)\right)^2}{1+\sum_{m=1}^{\infty}\delta_{*,m}^2}$$

$$
\begin{aligned}
&= \frac{\left(1+\sum_{m=1}^{\infty}\delta_m\rho_m(x)\right)^2-\left(1+\sum_{m=1}^{\infty}\delta_{*,m}\rho_m(x)\right)^2}{1+\sum_{m=1}^{\infty}\delta_m^2}\\
&\quad+\left(1+\sum_{m=1}^{\infty}\delta_{*,m}\rho_m(x)\right)^2\left(\frac{1}{1+\sum_{m=1}^{\infty}\delta_m^2}-\frac{1}{1+\sum_{m=1}^{\infty}\delta_{*,m}^2}\right)\\
&=\left(\sum_{m=1}^{\infty}(\delta_m-\delta_{*,m})\rho_m(x)\right)\left(\frac{2+\sum_{m=1}^{\infty}(\delta_m+\delta_{*,m})\rho_m(x)}{1+\sum_{m=1}^{\infty}\delta_m^2}\right)\\
&\quad+\left(1+\sum_{m=1}^{\infty}\delta_{*,m}\rho_m(x)\right)^2\left(\frac{1}{1+\sum_{m=1}^{\infty}\delta_m^2}-\frac{1}{1+\sum_{m=1}^{\infty}\delta_{*,m}^2}\right)\\
&=\left(\sum_{m=1}^{\infty}(\delta_m-\delta_{*,m})\rho_m(x)\right)\left(\frac{2+\sum_{m=1}^{\infty}(\delta_m+\delta_{*,m})\rho_m(x)}{1+\sum_{m=1}^{\infty}\delta_m^2}\right)\\
&\quad+\left(1+\sum_{m=1}^{\infty}\delta_{*,m}\rho_m(x)\right)^2\frac{\sum_{m=1}^{\infty}\left(\delta_{*,m}^2-\delta_m^2\right)}{\left(1+\sum_{m=1}^{\infty}\delta_m^2\right)\left(1+\sum_{m=1}^{\infty}\delta_{*,m}^2\right)}.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&\int_0^1 |f(x|\boldsymbol{\delta})-f(u|\boldsymbol{\delta}_*)|\,\mathrm{d}x\\
&\quad\le \frac{1}{1+\sum_{m=1}^{\infty}\delta_m^2}\int_0^1\left|\sum_{m=1}^{\infty}(\delta_m-\delta_{*,m})\rho_m(x)\right|\\
&\qquad\times\left|2+\sum_{m=1}^{\infty}(\delta_m+\delta_{*,m})\rho_m(x)\right|\mathrm{d}x\\
&\quad+\frac{\left|\sum_{m=1}^{\infty}(\delta_{*,m}^2-\delta_m^2)\right|}{\left(1+\sum_{m=1}^{\infty}\delta_m^2\right)\left(1+\sum_{m=1}^{\infty}\delta_{*,m}^2\right)}\times\int_0^1\left(1+\sum_{m=1}^{\infty}\delta_{*,m}\rho_m(x)\right)^2\mathrm{d}x\\
&\quad\le\frac{1}{1+\sum_{m=1}^{\infty}\delta_m^2}\sqrt{\int_0^1\left(\sum_{m=1}^{\infty}(\delta_m-\delta_{*,m})\rho_m(x)\right)^2\mathrm{d}x}\\
&\qquad\times\sqrt{\int_0^1\left(2+\sum_{m=1}^{\infty}(\delta_m+\delta_{*,m})\rho_m(x)\right)^2\mathrm{d}x}+\frac{\left|\sum_{m=1}^{\infty}(\delta_{*,m}^2-\delta_m^2)\right|}{1+\sum_{m=1}^{\infty}\delta_m^2}\\
&\quad=\sqrt{\sum_{m=1}^{\infty}(\delta_m-\delta_{*,m})^2}\times\frac{\sqrt{4+\sum_{m=1}^{\infty}(\delta_m+\delta_{*,m})^2}}{1+\sum_{m=1}^{\infty}\delta_m^2}+\frac{\left|\sum_{m=1}^{\infty}(\delta_{*,m}^2-\delta_m^2)\right|}{1+\sum_{m=1}^{\infty}\delta_m^2}
\end{aligned}
$$

$$= ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||. \times \frac{\sqrt{4 + ||\boldsymbol{\delta}_* + \boldsymbol{\delta}||^2}}{1 + ||\boldsymbol{\delta}||^2} + \frac{|\langle \boldsymbol{\delta}_* - \boldsymbol{\delta}, \boldsymbol{\delta}_* + \boldsymbol{\delta} \rangle|}{1 + ||\boldsymbol{\delta}||^2}, \tag{7.20}$$

where the second inequality follows from Schwarz inequality.

To evaluate the last expression in (7.20) further, note that

$$\begin{aligned} \frac{|\langle \boldsymbol{\delta}_* - \boldsymbol{\delta}, \boldsymbol{\delta}_* + \boldsymbol{\delta} \rangle|}{1 + ||\boldsymbol{\delta}||^2} &\leq \frac{||\boldsymbol{\delta}_* - \boldsymbol{\delta}||.||\boldsymbol{\delta}_* + \boldsymbol{\delta}||}{1 + ||\boldsymbol{\delta}||^2} \\ &\leq ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||.\frac{2||\boldsymbol{\delta}||}{1 + ||\boldsymbol{\delta}||^2} \\ &\leq ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||, \end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality and the last inequality follows from the trivial inequality $2||\boldsymbol{\delta}|| \leq 1 + ||\boldsymbol{\delta}||^2$. Moreover, note that

$$\begin{aligned} \frac{\sqrt{4 + ||\boldsymbol{\delta}_* + \boldsymbol{\delta}||^2}}{1 + ||\boldsymbol{\delta}||^2} &\leq \frac{\sqrt{4 + ||\boldsymbol{\delta}_* + \boldsymbol{\delta}||^2}}{\sqrt{1 + ||\boldsymbol{\delta}||^2}} \\ &\leq \sqrt{5 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||}, \end{aligned}$$

where the last inequality follows from

$$\begin{aligned} \frac{4 + (||\boldsymbol{\delta}_* + \boldsymbol{\delta}||)^2}{1 + ||\boldsymbol{\delta}||^2} &\leq \frac{4 + (||\boldsymbol{\delta}_* - \boldsymbol{\delta}|| + ||\boldsymbol{\delta}||)^2}{1 + ||\boldsymbol{\delta}||^2} \\ &= \frac{4 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + 2||\boldsymbol{\delta}||.||\boldsymbol{\delta}_* - \boldsymbol{\delta}|| + ||\boldsymbol{\delta}||^2}{1 + ||\boldsymbol{\delta}||^2} \\ &\leq 5 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||. \end{aligned}$$

Thus,

$$\begin{aligned} &\int_0^1 |f(x|\boldsymbol{\delta}) - f(x|\boldsymbol{\delta}_*)| \, \mathrm{d}x \\ &\leq ||\boldsymbol{\delta}_* - \boldsymbol{\delta}|| \sqrt{5 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||} + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||. \end{aligned}$$

In particular, if $||\boldsymbol{\delta}_* - \boldsymbol{\delta}|| < 1$ then $\sqrt{5 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 + ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||} < \sqrt{7} < 3$ and $||\boldsymbol{\delta}_* - \boldsymbol{\delta}||^2 \leq ||\boldsymbol{\delta}_* - \boldsymbol{\delta}||$, so that

$$\int_0^1 |f(x|\boldsymbol{\delta}) - f(x|\boldsymbol{\delta}_*)| \, \mathrm{d}x \leq 5.||\boldsymbol{\delta}_* - \boldsymbol{\delta}||.$$

# Chapter 8

# Uniform convergence of SNP functions

## 8.1 Introduction

As mentioned in Chapter 7, given an a priori chosen absolutely continuous distribution function $G(x)$ with support $\mathbb{R}$, any distribution function $F(x)$ with support $\mathbb{R}$ can be written as $F(x) = H(G(x))$, where $H(u) = F(G^{-1}(u))$ is a distribution function on $[0,1]$, with $G^{-1}(u)$ the inverse of $G$. Moreover, if $F$ is absolutely continuous with density $f$ then $H$ is absolutely continuous with density

$$h(u) = \frac{f(G^{-1}(u))}{g(G^{-1}(u))}, \tag{8.1}$$

where $g$ is the density of $G$. Thus,

$$f(x) = h(G(x)).g(x).$$

Since both $f(x)$ and $g(x)$ are positive on $\mathbb{R}$ it follows that $h(u) > 0$ on $(0,1)$, and if $f(x)$ and $g(x)$ are continuous on $\mathbb{R}$ then $h(u)$ is continuous on $(0,1)$ because $G^{-1}(u)$ is continuous on $(0,1)$. Therefore, given the continuity and positivity of $f(x)$ and $g(x)$ on $\mathbb{R}$, it follows from Theorem 7.4 that the density $h(u)$ in (8.1) has the unique cosine series representation

$$h(u) \quad = \quad h(u|\boldsymbol{\delta}) = \frac{\left(1 + \sum_{m=1}^{\infty} \delta_m \sqrt{2} \cos(m\pi u)\right)^2}{1 + \sum_{k=1}^{\infty} \delta_k^2} \text{ a.e. on } [0,1], \tag{8.2}$$

$$\text{with } \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} \text{ satisfying } \sum_{m=1}^{\infty} \delta_m^2 < \infty, \tag{8.3}$$

where by Theorem 7.5,

$$\delta_m = \frac{\int_0^1 \sqrt{2}\cos(m\pi u)\sqrt{h(u)}\mathrm{d}u}{\int_0^1 \sqrt{h(u)}\mathrm{d}u}, \quad m \in \mathbb{N}. \tag{8.4}$$

Recall that the corresponding SNP density takes the form

$$h_n(u) = h(u|\pi_n\boldsymbol{\delta}) = \frac{\left(1 + \sum_{m=1}^{n} \delta_m \sqrt{2}\cos(m\pi u)\right)^2}{1 + \sum_{k=1}^{n} \delta_k^2}, \tag{8.5}$$

where here and in the sequel $\pi_n$ is the truncation operator, i.e.,

**Definition 8.1**. *The operator $\pi_n$ applied to $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty}$ as $\pi_n\boldsymbol{\delta}$ replaces all the $\delta_m$'s for $m \geq n+1$ by zeros.*

It follows now from Theorem 7.4 that $\lim_{n\to\infty} h(u|\pi_n\boldsymbol{\delta}) = h(u)$ a.e. on $[0, 1]$. Thus, denoting

$$f(x|\pi_n\boldsymbol{\delta}) = h(G(x)|\pi_n\boldsymbol{\delta}).g(x),$$

we have $\lim_{n\to\infty} f(x|\pi_n\boldsymbol{\delta}) = f(x)$ a.e. on $\mathbb{R}$. Moreover, denoting

$$F(x|\pi_n\boldsymbol{\delta}) = H(G(x)|\pi_n\boldsymbol{\delta}),$$

with $H(u|\pi_n\boldsymbol{\delta})$ defined similar to $H_n(u)$ in Theorem 7.3, it follows from Theorem 7.4 that

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F(x|\pi_n\boldsymbol{\delta}) - F(x)| = 0.$$

In quite a few SNP models the non-Euclidean parameter involved is either a distribution function $F(x)$ or a density $f(x)$ on $\mathbb{R}$, or both, and therefore these models depend on an infinite dimensional parameter vector $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty}$. Moreover, quite a few of these models depend on a Euclidean parameter vector $\theta$ as well, which appears, via a parametric function, as argument of a unknown distribution and/or density function.

For example, the discrete choice model considered in Bierens (2014) takes the form $\Pr[Y = 1|X] = F_0(X'\theta_0)$, where $Y \in \{0, 1\}$ is the dichotomous dependent variable, $X \in \mathbb{R}^k$ is a vector of covariates, possibly including the constant 1, $\theta_0 \in \mathbb{R}^p$ is the Euclidean parameter vector and the distribution function $F_0$ is the non-Euclidean parameter. The latter can be modeled semi-nonparametrically as $F_0(x) = H(G(x)|\boldsymbol{\delta}^0)$, with $\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty}$, so that for some Euclidean parameter vector $\theta_0$ and an infinite-dimensional parameter $\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty}$,

$$\Pr[Y = 1|X] = H(G(X'\theta_0)|\boldsymbol{\delta}^0).$$

In Bierens (2014) I have shown that under some regularity and normalization conditions the parameters $\theta_0$ and $\boldsymbol{\delta}^0$ are identified and can be estimated consistently (given an appropriate metric) by sieve maximum likelihood (ML) estimation, on the basis of a random sample of size $N$ from $(Y, X)$. Moreover, in Bierens (2014) I have set forth conditions such that the sieve ML estimator $\widehat{\theta}_N$ of $\theta_0$ is asymptotically normal, $\sqrt{N}(\widehat{\theta}_N - \theta_0) \xrightarrow{\text{d}} N_p[0, \Sigma]$, and asymptotically efficient. The latter results, in the approach in Bierens (2014), require (among other regularity conditions) that $H(u|\boldsymbol{\delta})$ is twice differentiable in $u$ with $h(u|\boldsymbol{\delta})$ and its derivative $h'(u|\boldsymbol{\delta})$ to $u$ be uniformly continuous on $[0, 1]$, and that $h(u|\pi_n\boldsymbol{\delta}) \to h(u|\boldsymbol{\delta})$ and $h'(u|\pi_n\boldsymbol{\delta}) \to h'(u|\boldsymbol{\delta})$ sufficiently fast as $n \to \infty$. In Bierens (2014) I have imposed these conditions by *assuming* that for all $\boldsymbol{\delta}$ in an infinite dimensional parameter space $\Delta$, say, containing $\boldsymbol{\delta}^0$ in its interior,

$$\sum_{m=1}^{\infty} m^{\ell}|\delta_m| < \infty \text{ for some } \ell \in \mathbb{N}. \tag{8.6}$$

This implies that for all $\boldsymbol{\delta} \in \Delta$, $h(u|\boldsymbol{\delta})$ and its derivatives $h^{(i)}(u|\boldsymbol{\delta}) = \partial^i h(u|\boldsymbol{\delta})/(\partial u)^i$, $i = 1, 2, ...., \ell$, to $u$ are uniformly continuous on $[0, 1]$, hence

$$\sup_{0\leq u\leq 1} h(u|\boldsymbol{\delta}) < \infty, \ \max_{i=1,2,\ldots,\ell} \sup_{0\leq u\leq 1} |h^{(i)}(u|\boldsymbol{\delta})| < \infty, \tag{8.7}$$

and

$$\begin{array}{l} \lim_{n\to\infty} \sup_{0\leq u\leq 1} |h(u|\pi_n\boldsymbol{\delta}) - h(u|\boldsymbol{\delta})| = 0, \\ \lim_{n\to\infty} \max_{i=1,2,\ldots,\ell} \sup_{0\leq u\leq 1} |h^{(i)}(u|\pi_n\boldsymbol{\delta}) - h^{(i)}(u|\boldsymbol{\delta})| = 0. \end{array} \tag{8.8}$$

As argued in Bierens (2014), in general the necessary minimum value of $\ell$ depends on whether the likelihood function involves $h(u|\boldsymbol{\delta})$ as well, and whether the Euclidean parameter vector appears as an argument of $h(u|\boldsymbol{\delta})$. If so, the second derivative of the log-likelihood function to the Euclidean parameters involves $h^{(2)}(u|\boldsymbol{\delta})$, so that at least $\ell = 2$. This is the case for the SNP Tobit model in Chapter 9. However, as shown in Bierens (2014), a slightly larger $\ell$ than strictly necessary, say $\ell = 3$, is advantageous because it allows for a smaller sieve order.

In this chapter I will therefore derive conditions on $f$ and $G$ such that (8.6) holds for $\ell = 3$. Moreover, I will determine the rate of uniform convergence in (8.8).

## 8.2 The choice of $G$

As said before, for $h(u)$ to be bounded on $[0, 1]$ we must choose $G$ such that its density $g(x)$ has fatter, or at least not thinner, tails than $f(x)$. A well-known distribution on $\mathbb{R}$ with very fat tails is the standard Cauchy distribution, which has c.d.f.

$$G(x) = 0.5 + \pi^{-1} \arctan(x),$$

density $g(x) = \pi^{-1}(1 + x^2)^{-1}$ and inverse $G^{-1}(u) = \tan(\pi(u - 0.5))$. Then it follows from (8.1) that

$$h(u) = \pi \left(1 + \tan^2(\pi(u - 0.5))\right) f\left(\tan(\pi(u - 0.5))\right). \tag{8.9}$$

Clearly, $h(u)$ is bounded on $[0, 1]$ if $\sup_{x\in\mathbb{R}} x^2 f(x) < \infty$. Moreover, if $f(x)$ is continuous and positive on $\mathbb{R}$ then $h(u)$ is continuous and positive on $(0, 1)$, and if $\int_{-\infty}^{\infty} |x|f(x)\mathrm{d}x < \infty$ then

$$h(0) = \lim_{u\downarrow 0} h(u) = \lim_{x\to-\infty} x^2 f(x) = 0, \ \ h(1) = \lim_{u\uparrow 1} h(u) = \lim_{x\to\infty} x^2 f(x) = 0,$$

because $\int_{-\infty}^{\infty} |x|f(x)\mathrm{d}x < \infty$ implies $\lim_{|x|\to\infty} x^2 f(x) = 0$. The latter follows from the following more general lemma, with $\psi(x) = xf(x)$.

**Lemma 8.1.** *Let $\psi(x)$ be a continuous real function on $\mathbb{R}$ such that the set $\{x \in \mathbb{R} : \psi(x) = 0\}$ is either finite or empty, and $\int_{-\infty}^{\infty} |\psi(x)|\mathrm{d}x < \infty$. Then $\lim_{|x|\to\infty} x\psi(x) = 0$.*

Thus, the following result holds.

**Theorem 8.1**. *Let $f(x)$ be a continuous and positive density on $\mathbb{R}$ satisfying $\int_{-\infty}^{\infty} |x|f(x)\mathrm{d}x < \infty$. Choose for $G$ the c.d.f. of the standard Cauchy distribution. Then the density $h(u)$ in (8.1) takes the form (8.9), and is uniformly continuous on $[0, 1]$ with tails $h(0) = h(1) = 0$.*

Given the choice of the standard Cauchy distribution function for $G$, I will now set forth additional conditions on the density $f$ such that (8.6) holds for $\ell = 3$, by deriving the cosine series representation of $\varphi(u) = \sqrt{h(u)}$ via integrating the cosine series representations of a higher-order derivative of $\varphi(u)$.

## 8.3 Uniform convergence by integration

In first instance, let $\varphi(u)$ be a twice differentiable real function on $(0,1)$, with first and second derivatives $\varphi'(u)$ and $\varphi''(u)$, respectively, such that $\varphi''(u) \in L^2(0,1)$. Denote

$$\begin{aligned}
\alpha_0 &= \int_0^1 \varphi''(u)\mathrm{d}u, \\
\alpha_m &= \int_0^1 \varphi''(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u \text{ for } m \in \mathbb{N}, \\
\varphi''_n(u) &= \alpha_0 + \sum_{m=1}^{n} \alpha_m \sqrt{2}\cos(m\pi u) \text{ for } n \in \mathbb{N}.
\end{aligned}$$

Then by Theorem 6.1,

$$\lim_{n\to\infty} \int_0^1 \left(\varphi''(u) - \varphi''_n(u)\right)^2 \mathrm{d}u = \lim_{n\to\infty} \sum_{m=n+1}^{\infty} \alpha_m^2 = 0, \tag{8.10}$$

where the last equality follows from

$$\int_0^1 \varphi''(u)^2 \mathrm{d}u = \sum_{m=0}^{\infty} \alpha_m^2 < \infty, \tag{8.11}$$

Recall that (8.10) implies $\varphi''(u) = \alpha_0 + \sum_{m=1}^{\infty} \alpha_m \sqrt{2}\cos(m\pi u)$ a.e. on $(0,1)$. Moreover, note that $\alpha_0 = \int_0^1 \varphi''(u)\mathrm{d}u = \varphi'(1) - \varphi'(0)$ and that by (8.11), $\alpha_0^2 < \infty$, so that we must have

$$\left(\varphi'(1) - \varphi'(0)\right)^2 < \infty. \tag{8.12}$$

More generally, for any $u \in (0,1)$ we have

$$|\varphi'(u) - \varphi'(0)| = \left|\int_0^u \varphi''(v)\mathrm{d}v\right| \leq \int_0^1 |\varphi''(v)|\mathrm{d}v \leq \sqrt{\int_0^1 \varphi''(v)^2\mathrm{d}v} < \infty. \tag{8.13}$$

Since $\varphi'(u)$ is differentiable and therefore continuous on $(0,1)$, it is uniformly continuous on any closed interval in $(0,1)$, so that $|\varphi'(u)| < \infty$ for each $u \in (0,1)$. Thus, it follows from (8.13) that $|\varphi'(0)| < \infty$, which by (8.12) implies that $|\varphi'(1)| < \infty$. Therefore, interpreting $\varphi'(0) = \lim_{u\downarrow 0}\varphi'(u)$ and $\varphi'(1) = \lim_{u\uparrow 1}\varphi'(u)$ it follows that $\varphi'(u)$ is uniformly continuous on $[0,1]$, so that $\sup_{0\leq u\leq 1}|\varphi'(u)| < \infty$. The latter implies that $\varphi'(u) \in L^2(0,1)$. By the same argument it follows that $\varphi(u)$ is uniformly continuous on $[0,1]$.

Now the primitive of $\varphi_n''(u)$ takes the general form

$$\varphi_n'(u) = c + (\varphi'(1) - \varphi'(0))u + \sum_{m=1}^{n} \frac{\alpha_m}{\pi m}\sqrt{2}\sin(m\pi u)$$

for some constant $c$. Since $\varphi_n'(1) = \varphi'(1)$ and $\varphi_n'(0) = \varphi'(0)$ if $c = \varphi'(0)$, the latter is a natural choice for $c$, so that

$$\begin{aligned}\varphi_n'(u) &= \varphi'(0) + (\varphi'(1) - \varphi'(0))u + \sum_{m=1}^{n} \frac{\alpha_m}{\pi m}\sqrt{2}\sin(m\pi u) \\ &= \varphi'(0) + \int_0^u \varphi_n''(v)\mathrm{d}v.\end{aligned}$$

Since also $\varphi'(u) = \varphi'(0) + \int_0^u \varphi''(v)\mathrm{d}v$ we have

$$\begin{aligned}\limsup_{n\to\infty}\sup_{0\leq u\leq 1}|\varphi'(u) - \varphi_n'(u)| &\leq \limsup_{n\to\infty}\int_0^1 |\varphi''(v) - \varphi_n''(v)|\,\mathrm{d}v \\ &\leq \limsup_{n\to\infty}\sqrt{\int_0^1 \left(\varphi''(v) - \varphi_n''(v)\right)^2\mathrm{d}v} \\ &= \sqrt{\lim_{n\to\infty}\sum_{m=n+1}^{\infty}\alpha_m^2} = 0.\end{aligned} \tag{8.14}$$

Consequently, the series expansion

$$\varphi'(u) \equiv \varphi'(0) + (\varphi'(1) - \varphi'(0))u + \sum_{m=1}^{\infty} \frac{\alpha_m}{\pi m}\sqrt{2}\sin(m\pi u) \tag{8.15}$$

holds exactly and uniformly on $[0,1]$.

However, a stronger result than (8.14) applies, namely

$$\sup_{0\leq u\leq 1}|\varphi'(u)-\varphi'_n(u)|=o(n^{-1/2}),$$

due to the following lemma.

**Lemma 8.2**. *$\sum_{m=1}^{\infty}\alpha_m^2<\infty$ implies that for $c>1/2$, $\sum_{m=n+1}^{\infty}m^{-c}|\alpha_m|=o\left(n^{1/2-c}\right)$.*

Next, the primitive $\varphi(u)$ of $\varphi'(u)$ takes the general form

$$\begin{aligned}\varphi(u) &= c+\varphi'(0)u+\frac{1}{2}(\varphi'(1)-\varphi'(0))u^2\\ &\quad -\sum_{m=1}^{\infty}\frac{\alpha_m}{(\pi m)^2}\sqrt{2}\cos(m\pi u)\end{aligned}\tag{8.16}$$

for some constant $c$. Since I have already shown that $\varphi(u)$ is uniformly continuous on $[0,1]$ and therefore $\sup_{0\leq u\leq 1}|\varphi(u)|<\infty$, it follows that $\varphi\in L^2(0,1)$. Thus by Theorem 6.1, $\varphi(u)$ has also the cosine series representation

$$\varphi(u)=\gamma_0+\sum_{m=1}^{\infty}\gamma_m\sqrt{2}\cos(m\pi u)\text{ a.e. on }(0,1),\tag{8.17}$$

where

$$\gamma_0=\int_0^1\varphi(u)\mathrm{d}u,\ \gamma_m=\int_0^1\varphi(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u\text{ for }m\in\mathbb{N}.$$

Now observe that for $m\in\mathbb{N}$,

$$\begin{aligned}\int_0^1 u\sqrt{2}\cos(m\pi u)\mathrm{d}u &= \sqrt{2}\frac{(-1)^m-1}{(m\pi)^2},\\ \int_0^1 u^2\sqrt{2}\cos(m\pi u)\mathrm{d}u &= \frac{2\sqrt{2}(-1)^m}{(m\pi)^2},\end{aligned}$$

hence by Theorem 6.1 and $\lim_{n\to\infty}\sum_{m=n+1}^{\infty}m^{-2}=0$ it follows that

$$u\equiv\frac{1}{2}+\sqrt{2}\sum_{m=1}^{\infty}\frac{(-1)^m-1}{(m\pi)^2}\sqrt{2}\cos(m\pi u),\tag{8.18}$$

$$u^2\equiv\frac{1}{3}+2\sqrt{2}\sum_{m=1}^{\infty}\frac{(-1)^m}{(m\pi)^2}\sqrt{2}\cos(m\pi u),\tag{8.19}$$

exactly and uniformly on $[0, 1]$. Therefore, (8.16) reads

$$\begin{aligned}\varphi(u) &= c + \frac{1}{2}\varphi'(0) + \frac{1}{6}\left(\varphi'(1) - \varphi'(0)\right) \\ &\quad + \sqrt{2}\varphi'(0)\sum_{m=1}^{\infty}\frac{(-1)^m - 1}{(m\pi)^2}\sqrt{2}\cos(m\pi u) \\ &\quad + \sqrt{2}\left(\varphi'(1) - \varphi'(0)\right)\sum_{m=1}^{\infty}\frac{(-1)^m}{(m\pi)^2}\sqrt{2}\cos(m\pi u) \\ &\quad - \sum_{m=1}^{\infty}\frac{\alpha_m}{(\pi m)^2}\sqrt{2}\cos(m\pi u) \\ &= c + \frac{1}{2}\varphi'(0) + \frac{1}{6}\left(\varphi'(1) - \varphi'(0)\right) \\ &\quad + \frac{\sqrt{2}\left(\varphi'(1)(-1)^m - \varphi'(0)\right) - \alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u)\end{aligned}$$

Thus, we must have that $\gamma_0 = \int_0^1 \varphi(u)\mathrm{d}u = c + \frac{1}{2}\varphi'(0) + \frac{1}{6}(\varphi'(1) - \varphi'(0))$, hence

$$c = \int_0^1 \varphi(u)\mathrm{d}u - \frac{1}{2}\varphi'(0) - \frac{1}{6}\left(\varphi'(1) - \varphi'(0)\right),$$

and for $m \in \mathbb{N}$,

$$\begin{aligned}\gamma_m &= \int_0^1 \varphi(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u \\ &= \frac{\sqrt{2}\left(\varphi'(1)(-1)^m - \varphi'(0)\right) - \alpha_m}{(m\pi)^2} \\ &= \frac{\sqrt{2}\left(\varphi'(1)(-1)^m - \varphi'(0)\right)}{(m\pi)^2} \\ &\quad - \frac{\int_0^1 \varphi''(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u}{(m\pi)^2}.\end{aligned} \tag{8.20}$$

Therefore, $\varphi(u)$ has two equivalent cosine series representations, namely

$$\begin{aligned}\varphi(u) &\equiv \int_0^1 \varphi(v)\mathrm{d}v - \frac{1}{2}\varphi'(0) - \frac{1}{6}\left(\varphi'(1) - \varphi'(0)\right) \\ &\quad + \varphi'(0)u + \frac{1}{2}(\varphi'(1) - \varphi'(0))u^2\end{aligned}$$

$$
\begin{aligned}
& -\sum_{m=1}^{\infty} \frac{\alpha_m}{(\pi m)^2} \sqrt{2}\cos(m\pi u) \\
\equiv & \int_0^1 \varphi(v)\mathrm{d}v + \sum_{m=1}^{\infty} \gamma_m \sqrt{2}\cos(m\pi u),
\end{aligned}
$$

exactly and uniformly on $[0,1]$. Moreover, denoting

$$
\begin{aligned}
\varphi_n(u) \equiv & \int_0^1 \varphi(v)\mathrm{d}v - \frac{1}{2}\varphi'(0) - \frac{1}{6}\left(\varphi'(1) - \varphi'(0)\right) \\
& + \varphi'(0)u + \frac{1}{2}(\varphi'(1) - \varphi'(0))u^2 \\
& - \sum_{m=1}^{n} \frac{\alpha_m}{(\pi m)^2} \sqrt{2}\cos(m\pi u)
\end{aligned}
$$

it follows from Lemma 8.2 that

$$
\sup_{0 \le u \le 1} |\varphi(u) - \varphi_n(u)| = o(n^{-3/2}).
$$

Furthermore, denoting

$$
\varphi_n^*(u) = \int_0^1 \varphi(v)\mathrm{d}v + \sum_{m=1}^{\infty} \gamma_m \sqrt{2}\cos(m\pi u),
$$

it follows from (8.20), Lemma 8.1 and the easy inequality $\sum_{m=n+1}^{\infty} m^{-2} \le \int_n^{\infty} x^{-2}\mathrm{d}x = n^{-1}$ that

$$
\begin{aligned}
\sup_{0 \le u \le 1} |\varphi(u) - \varphi_n^*(u)| \le & \frac{2}{\pi^2}\left(|\varphi'(1)| + |\varphi'(0)|\right) \sum_{m=n+1}^{\infty} m^{-2} \\
& + \frac{\sqrt{2}}{\pi^2} \sum_{m=n+1}^{\infty} m^{-2}|\alpha_m| \\
= & O(n^{-1}) + o(n^{-3/2}) = O(n^{-1}).
\end{aligned}
$$

Summarizing, the following results hold.

**Theorem 8.2**. *Suppose that $\varphi(u)$ is a twice differentiable real function on $(0,1)$, with first and second derivatives $\varphi'(u)$ and $\varphi''(u)$, respectively, and* *that* $\varphi'' \in L^2(0,1)$. *Then* $\varphi(u)$ *and* $\varphi'(u)$ *are uniformly continuous on* $[0,1]$. *Next, denote*

$$\begin{aligned}\alpha_0 &= \int_0^1 \varphi''(u)\mathrm{d}u,\ \alpha_m = \int_0^1 \varphi''(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u\ for\ m \in \mathbb{N},\\ \gamma_0 &= \int_0^1 \varphi(u)\mathrm{d}u,\ \gamma_m = \int_0^1 \varphi(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u\ for\ \ m \in \mathbb{N}.\end{aligned}$$

*Then* $\varphi(u)$ *and* $\varphi'(u)$ *have the cosine-sine series representations*

$$\varphi'(u) \equiv \varphi'(0) + (\varphi'(1) - \varphi'(0))\,u + \sum_{m=1}^{\infty} \frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u), \quad (8.21)$$

$$\varphi(u) \equiv \gamma_0 + \sum_{m=1}^{\infty} \gamma_m \sqrt{2}\cos(m\pi u) \quad (8.22)$$

$$\begin{aligned}&\equiv \int_0^1 \varphi(v)\mathrm{d}v - \frac{1}{2}\varphi'(0) - \frac{1}{6}(\varphi'(1) - \varphi'(0))\\ &\quad +\varphi'(0)u + \frac{1}{2}(\varphi'(1) - \varphi'(0))\,u^2\\ &\quad -\sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u), \quad (8.23)\end{aligned}$$

*exactly and uniformly on* $[0,1]$, *where for* $m \in \mathbb{N}$,

$$\gamma_m = \frac{\sqrt{2}\,(\varphi'(1)(-1)^m - \varphi'(0)) - \alpha_m}{(m\pi)^2}. \quad (8.24)$$

*Consequently, denoting*

$$\varphi_n'(u) = \varphi'(0) + (\varphi'(1) - \varphi'(0))\,u + \sum_{m=1}^{n} \frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u)$$

*and*

$$\begin{aligned}\varphi_n(u) &= \int_0^1 \varphi(v)\mathrm{d}v - \frac{1}{2}\varphi'(0) - \frac{1}{6}(\varphi'(1) - \varphi'(0))\\ &\quad +\varphi'(0)u + \frac{1}{2}(\varphi'(1) - \varphi'(0))\,u^2\\ &\quad -\sum_{m=1}^{n} \frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u) \quad (8.25)\end{aligned}$$

*we have*

$$\sup_{0\leq u\leq 1}|\varphi'(u)-\varphi_n'(u)| = o(n^{-1/2}),\ \sup_{0\leq u\leq 1}|\varphi(u)-\varphi_n(u)| = o(n^{-3/2}). \quad (8.26)$$

*Moreover, defining*

$$\varphi_n^*(u) = \gamma_0 + \sum_{m=1}^{n}\gamma_m\sqrt{2}\cos(m\pi u) \quad (8.27)$$

*instead of* (8.25), *it follows that*

$$\sup_{0\leq u\leq 1}|\varphi(u)-\varphi_n^*(u)| = O(1/n). \quad (8.28)$$

As said before, in some SNP models, the Euclidean parameters enter the model via the argument of an unknown real function $\varphi(u)$ on $[0,1]$, in particular $\varphi(u) = \sqrt{h(u)}$ with $h(u)$ a density, and deriving the asymptotic normality of the sieve estimators of these Euclidean parameters around their true values along the lines in Bierens (2014) requires that the first and second derivatives of $\varphi(u)$ are "smooth" and the parameters of their sine-cosine series representations converge to zero at a fast enough rate.

To enforce these conditions, suppose that now that $\varphi(u)$ is a four times differentiable real function on $(0,1)$, with fourth derivative $\varphi'''' \in L^2(0,1)$. Then it follows from Theorem 8.2 that

$$\begin{aligned}
\varphi'''(u) &\equiv \varphi'''(0) + (\varphi'''(1)-\varphi'''(0))\,u + \sum_{m=1}^{\infty}\frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u),\\
\varphi''(u) &\equiv \eta_0 + \sum_{m=1}^{\infty}\eta_m\sqrt{2}\cos(m\pi u)\mathrm{d}u\\
&\equiv \int_0^1\varphi''(v)\mathrm{d}v - \frac{1}{2}\varphi'''(0) - \frac{1}{6}(\varphi'''(1)-\varphi'''(0))\\
&\quad +\varphi'''(0)u + \frac{1}{2}(\varphi'''(1)-\varphi'''(0))\,u^2 - \sum_{m=1}^{\infty}\frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u)\\
&\equiv (\varphi'(1)-\varphi'(0)) - \frac{1}{2}\varphi'''(0) - \frac{1}{6}(\varphi'''(1)-\varphi'''(0))\\
&\quad +\varphi'''(0)u + \frac{1}{2}(\varphi'''(1)-\varphi'''(0))\,u^2 - \sum_{m=1}^{\infty}\frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u)
\end{aligned}$$

exactly and uniformly on $[0,1]$, where now

$$\begin{aligned}
\alpha_0 &= \int_0^1 \varphi''''(u)\mathrm{d}u = \varphi'''(1) - \varphi'''(0) \\
\alpha_m &= \int_0^1 \varphi''''(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u \text{ for } m \in \mathbb{N}, \\
\eta_0 &= \int_0^1 \varphi''(u)\mathrm{d}u = \varphi'(1) - \varphi'(0), \\
\eta_m &= \int_0^1 \varphi''(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u. \qquad (8.29)
\end{aligned}$$

At this point we can write

$$\begin{aligned}
\varphi'''(u) &\equiv P_1(u) + \sum_{m=1}^{\infty} \frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u), \\
\varphi''(u) &\equiv P_2(u) - \sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u), \\
\varphi'(u) &\equiv P_3(u) - \sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^3}\sqrt{2}\sin(m\pi u), \qquad (8.30) \\
\varphi(u) &\equiv P_4(u) + \sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^4}\sqrt{2}\cos(m\pi u),
\end{aligned}$$

where

$$\begin{aligned}
P_1(u) &= \varphi'''(0) + (\varphi'''(1) - \varphi'''(0))\,u, \qquad (8.31) \\
P_2(u) &= \varphi'(1) - \varphi'(0) - \frac{1}{2}\varphi'''(0) - \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right) \\
&\quad + \varphi'''(0)u + \frac{1}{2}\left(\varphi'''(1) - \varphi'''(0)\right)u^2 \qquad (8.32)
\end{aligned}$$

and

$$\begin{aligned}
P_3(u) &= c_3 + \int_0^u P_2(v)\mathrm{d}v \\
&= c_3 + \left(\varphi'(1) - \varphi'(0)\right)u - \frac{1}{2}\varphi'''(0)u - \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right)u \\
&\quad + \frac{1}{2}\varphi'''(0)u^2 + \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right)u^3 \\
P_4(u) &= c_4 + \int_0^u P_3(v)\mathrm{d}v
\end{aligned}$$

for some constants $c_3$ and $c_4$.

As to the choice of $c_3$, note that $P_3(0) = c_3$, $P_3(1) = c_3 + \varphi'(1) - \varphi'(0)$, hence if $c_3 = \varphi'(0)$ then (8.30) fits exactly on $[0,1]$. Thus,

$$\begin{aligned} P_3(u) &= \varphi'(0) + \int_0^u P_2(v)\mathrm{d}v \\ &= \varphi'(0) + \left(\varphi'(1) - \varphi'(0) - \frac{1}{2}\varphi'''(0) - \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right)\right) u \\ &\quad + \frac{1}{2}\varphi'''(0)u^2 + \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right) u^3 \end{aligned} \tag{8.33}$$

so that

$$\begin{aligned} P_4(u) &= c_4 + \int_0^u P_3(v)\mathrm{d}v \\ &= c_4 + \varphi'(0).u \\ &\quad + \frac{1}{2}\left(\varphi'(1) - \varphi'(0) - \frac{1}{2}\varphi'''(0) - \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right)\right) u^2 \\ &\quad + \frac{1}{6}\varphi'''(0)u^3 + \frac{1}{24}\left(\varphi'''(1) - \varphi'''(0)\right) u^4 \end{aligned}$$

for some constant $c_4$.

As before, we can also write

$$\varphi(u) = \gamma_0 + \sum_{m=1}^{\infty} \gamma_m \sqrt{2}\cos(m\pi u) \text{ a.e. on } (0,1)$$

where

$$\gamma_0 = \int_0^1 \varphi(u)\mathrm{d}u,\ \gamma_m = \int_0^1 \varphi(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u \text{ for } m \in \mathbb{N}.$$

Therefore,

$$\gamma_0 = \int_0^1 \varphi(u)\mathrm{d}u = \int_0^1 P_4(u)\mathrm{d}u,$$

hence

$$\begin{aligned} c_4 &= \int_0^1 \varphi(v)\mathrm{d}v - \frac{1}{2}\varphi'(0) \\ &\quad - \frac{1}{6}\left(\varphi'(1) - \varphi'(0) - \frac{1}{2}\varphi'''(0) - \frac{1}{6}\left(\varphi'''(1) - \varphi'''(0)\right)\right) \\ &\quad - \frac{1}{24}\varphi'''(0) - \frac{1}{120}\left(\varphi'''(1) - \varphi'''(0)\right), \end{aligned}$$

so that

$$\begin{aligned}
P_4(u) &= \int_0^1 \varphi(v)\mathrm{d}v - \frac{1}{2}\varphi'(0) \\
&\quad -\frac{1}{6}\left(\varphi'(1)-\varphi'(0)-\frac{1}{2}\varphi'''(0)-\frac{1}{6}\left(\varphi'''(1)-\varphi'''(0)\right)\right) \\
&\quad -\frac{1}{24}\varphi'''(0)-\frac{1}{120}\left(\varphi'''(1)-\varphi'''(0)\right)+\varphi'(0).u \\
&\quad +\frac{1}{2}\left(\varphi'(1)-\varphi'(0)-\frac{1}{2}\varphi'''(0)-\frac{1}{6}\left(\varphi'''(1)-\varphi'''(0)\right)\right)u^2 \\
&\quad +\frac{1}{6}\varphi'''(0)u^3+\frac{1}{24}\left(\varphi'''(1)-\varphi'''(0)\right)u^4, \qquad (8.34)
\end{aligned}$$

and

$$\gamma_m = \int_0^1 P_4(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u + \frac{\alpha_m}{(m\pi)^4},\ \ m \in \mathbb{N}.$$

Finally, note that similar to Theorem 8.2 the left and right tails of $\varphi(u)$, $\varphi'(u)$, $\varphi''(u)$ and $\varphi'''(u)$ are all finite.

Summarizing, the following results hold.

**Theorem 8.3**. *Suppose that $\varphi(u)$ is a four times differentiable real function on $(0,1)$ with fourth derivative $\varphi'''' \in L^2(0,1)$. Denote for $m \in \mathbb{N}$,*

$$\alpha_m = \int_0^1 \varphi''''(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u.$$

*Then there exists a polynomial $P_4(u)$ of order 4, given by (8.34), such that exactly and uniformly on $[0,1]$,*

$$\begin{aligned}
\varphi(u) &\equiv P_4(u) + \sum_{m=1}^{\infty}\frac{\alpha_m}{(m\pi)^4}\sqrt{2}\cos(m\pi u), \\
\varphi'(u) &\equiv P_4'(u) - \sum_{m=1}^{\infty}\frac{\alpha_m}{(m\pi)^3}\sqrt{2}\sin(m\pi u), \\
\varphi''(u) &\equiv P_4''(u) - \sum_{m=1}^{\infty}\frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u), \\
\varphi'''(u) &\equiv P_4'''(u) + \sum_{m=1}^{\infty}\frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u).
\end{aligned}$$

*Moreover, denoting for $n \in \mathbb{N}$,*

$$\begin{aligned}
\varphi_n(u) &= P_4(u) + \sum_{m=1}^{n} \frac{\alpha_m}{(m\pi)^4}\sqrt{2}\cos(m\pi u), \\
\varphi_n'(u) &= P_4'(u) - \sum_{m=1}^{n} \frac{\alpha_m}{(m\pi)^3}\sqrt{2}\sin(m\pi u), \\
\varphi_n''(u) &= P_4''(u) - \sum_{m=1}^{n} \frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u), \\
\varphi_n'''(u) &= P_4'''(u) + \sum_{m=1}^{n} \frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u),
\end{aligned}$$

*it follows from Lemma 8.1 that*

$$\left.\begin{array}{l}
\sup_{0\le u\le 1} |\varphi(u) - \varphi_n(u)| = o(n^{-7/2}), \\
\sup_{0\le u\le 1} |\varphi'(u) - \varphi_n'(u)| = o(n^{-5/2}), \\
\sup_{0\le u\le 1} |\varphi''(u) - \varphi_n''(u)| = o(n^{-3/2}), \\
\sup_{0\le u\le 1} |\varphi'''(u) - \varphi_n'''(u)| = o(n^{-1/2}).
\end{array}\right\} \quad (8.35)$$

**Remark**. Note that if $\varphi'''(1) = \varphi'''(0) = 0$ and $\varphi'(1) = \varphi'(0) = 0$ then by (8.31), (8.32), (8.33) and (8.34), $P_1(u) = P_2(u) = P_3(u) = 0$ and $P_4(u) = \int_0^1 \varphi(v)\mathrm{d}v$, so that exactly and uniformly on $[0,1]$,

$$\begin{aligned}
\varphi(u) &\equiv \int_0^1 \varphi(v)\mathrm{d}v + \sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^4}\sqrt{2}\cos(m\pi u), \\
\varphi'(u) &\equiv -\sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^3}\sqrt{2}\sin(m\pi u), \\
\varphi''(u) &\equiv -\sum_{m=1}^{\infty} \frac{\alpha_m}{(m\pi)^2}\sqrt{2}\cos(m\pi u), \\
\varphi'''(u) &\equiv \sum_{m=1}^{\infty} \frac{\alpha_m}{m\pi}\sqrt{2}\sin(m\pi u),
\end{aligned}$$

and now

$$\gamma_m = \int_0^1 \varphi(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u = \frac{\alpha_m}{(m\pi)^4}, \; m \in \mathbb{N}.$$

Thus,

$$\sum_{m=1}^{\infty} m^8 \gamma_m^2 = \pi^{-8} \sum_{m=1}^{\infty} \alpha_m^2 < \infty, \tag{8.36}$$

which by Lemma 8.1, with $\alpha_m$ replaced by $m^4\gamma_m$, implies that

$$\begin{array}{l} \sum_{m=n+1}^{\infty} m^3|\gamma_m| = \sum_{m=n+1}^{\infty} m^{-1}|m^4\gamma_m| = o(n^{-1/2}), \\ \sum_{m=n+1}^{\infty} m^2|\gamma_m| = \sum_{m=n+1}^{\infty} m^{-2}|m^4\gamma_m| = o(n^{-3/2}), \\ \sum_{m=n+1}^{\infty} m|\gamma_m| = \sum_{m=n+1}^{\infty} m^{-3}|m^4\gamma_m| = o(n^{-5/2}), \\ \sum_{m=n+1}^{\infty} |\gamma_m| = \sum_{m=n+1}^{\infty} m^{-4}|m^4\gamma_m| = o(n^{-7/2}). \end{array} \tag{8.37}$$

Note that the first result in (8.37) implies that $\sum_{m=1}^{\infty} m^3|\gamma_m| < \infty$.

Summarizing, the following results holds.

**Theorem 8.4**. *Let $\varphi(u)$ be a four times differentiable real function on $(0,1)$, with fourth derivative $\varphi''''(u) \in L^2(0,1)$, and satisfying the tail conditions*

$$\lim_{u\downarrow 0} \varphi'''(u) = \lim_{u\uparrow 1} \varphi'''(u) = 0, \ \lim_{u\downarrow 0} \varphi'(u) = \lim_{u\uparrow 1} \varphi'(u) = 0.$$

*Then $\varphi(u)$ and its first, second and third derivatives are uniform continuous on $[0,1]$ and have the exact and uniform cosine-sine series representations*

$$\begin{aligned} \varphi(u) &\equiv \int_0^1 \varphi(u)\mathrm{d}u + \sum_{m=1}^{\infty} \gamma_m \sqrt{2}\cos(m\pi u), & (8.38) \\ \varphi'(u) &\equiv -\pi \sum_{m=1}^{\infty} m.\gamma_m \sqrt{2}\sin(m\pi u), \\ \varphi''(u) &\equiv -\pi^2 \sum_{m=1}^{\infty} m^2\gamma_m \sqrt{2}\cos(m\pi u), \\ \varphi'''(u) &\equiv \pi^3 \sum_{m=1}^{\infty} m^3\gamma_m \sqrt{2}\sin(m\pi u) \end{aligned}$$

*on $[0,1]$, where $\gamma_m = \int_0^1 \varphi(u)\sqrt{2}\cos(m\pi u)\mathrm{d}u$ for $m \in \mathbb{N}$, satisfying $\sum_{m=n+1}^{\infty} m^3|\gamma_m| = o(n^{-1/2})$, hence*

$$\sum_{m=1}^{\infty} m^3|\gamma_m| < \infty. \tag{8.39}$$

*Moreover, denoting* $\varphi_n(u) = \int_0^1 \varphi(u)\mathrm{d}u + \sum_{m=1}^{n} \gamma_m \sqrt{2}\cos(m\pi u)$, *the uniform convergence rates* (8.35) *apply.*

**Remark**. Note that conditions in Theorem 8.4 generate a *Sobolev space* of functions $\varphi(u)$ on $[0,1]$. See for example Adams and Fournier (2003).

Now suppose that in Theorem 8.4, $\varphi(u) = \sqrt{h(u)}$, where $h(u)$ is a positive density on $(0,1)$, so that by (8.2),

$$\varphi(u) = \sqrt{h(u)} = \frac{1 + \sum_{m=1}^{\infty} \delta_m \sqrt{2}\cos(m\pi u)}{\sqrt{1 + \sum_{k=1}^{\infty} \delta_k^2}},$$

where by (8.4) and (8.38), $\delta_m = \left(\int_0^1 \varphi(u)\mathrm{d}u\right)^{-1} \gamma_m$, so that by (8.39), $\sum_{m=1}^{\infty} m^3|\delta_m| < \infty$. Moreover, it follows from (8.36) that

$$\sum_{k=n+1}^{\infty} \delta_k^2 = \sum_{k=n+1}^{\infty} m^{-8} m^8 \delta_k^2 \leq \sum_{k=n+1}^{\infty} m^{-8} \sum_{k=n+1}^{\infty} m^8 \delta_k^2 = o(n^{-7}).$$

It is now easy to verify from Theorem 8.4 that the following results hold.

**Theorem 8.5**. *Let* $h(u)$ *be a positive density on* $(0,1)$ *such that* $\varphi(u) = \sqrt{h(u)}$ *satisfies the conditions of Theorem 8.4.*[1] *Then* $h(u)$ *has exactly and uniformly the cosine series representation* (8.2), *where the* $\delta_m$*'s are unique and satisfy*

$$\sum_{m=1}^{\infty} m^3|\delta_m| < \infty. \tag{8.40}$$

*Moreover, with* $h_n(u)$ *defined by* (8.5), *the following uniform convergence rates apply*:

$$\begin{aligned}
&\sup_{0\leq u\leq 1} |h(u) - h_n(u)| = o(n^{-7/2}),\\
&\sup_{0\leq u\leq 1} |h'(u) - h_n'(u)| = o(n^{-5/2}),\\
&\sup_{0\leq u\leq 1} |h''(u) - h_n''(u)| = o(n^{-3/2}),\\
&\sup_{0\leq u\leq 1} |h'''(u) - h_n'''(u)| = o(n^{-1/2}).
\end{aligned}$$

[1] So that $h(u)$ is uniformly continuous on $[0,1]$.

## 8.4 Moment conditions

The next question I will address is: Under what conditions on the density $f(x)$ are the conditions of Theorem 8.5 satisfied for $h(u)$ defined by (8.9)?

In the latter case,

$$\varphi(u) = \sqrt{h(u)} = \sqrt{\pi}\left(\sqrt{1+\tan^2(\pi(u-0.5))}\right)\eta\left(\tan(\pi(u-0.5))\right), \quad (8.41)$$

where for notational convenience,

$$\eta(x) = \sqrt{f(x).}$$

Of course, in order that $\varphi(u)$ is four times differentiable on $(0,1)$ we need to require that $\eta(x)$ is four times differentiable on $\mathbb{R}$, hence $f(x)$ needs to be four times differentiable on $\mathbb{R}$.

Now denote

$$\begin{aligned}\phi(u|k,c,\psi) &= \tan^k(\pi(u-0.5))\left(1+\tan^2(\pi(u-0.5))\right)^c \\ &\quad \times\psi\left(\tan(\pi(u-0.5))\right) \qquad (8.42)\end{aligned}$$

for some differentiable real function $\psi$ on $\mathbb{R}$, so that

$$\varphi(u) = \sqrt{\pi}\phi(u|0,1/2,\eta), \quad (8.43)$$

Using the well-known fact that

$$\mathrm{d}\tan(\pi(u-0.5))/\mathrm{d}u = \pi\left(1+\tan^2(\pi(u-0.5))\right),$$

it follows that

$$\begin{aligned}\phi'(u|k,c,\psi) &= \partial\phi(u|k,c,\psi)/\partial u \\ &= k\pi\phi(u|k-1,c+1,\psi) + 2c\pi\phi(u|k+1,c,\psi) \\ &\quad +\pi\phi(u|k,c+1,\psi'), \qquad (8.44)\end{aligned}$$

hence, plugging in $k=0$, $c=0.5$ and $\psi=\eta$, it follows from (8.44) that

$$\varphi'(u) = \pi\sqrt{\pi}\phi(u|1,0.5,\eta) + \pi\sqrt{\pi}\phi(u|0,1.5,\eta'). \quad (8.45)$$

Similarly, it follows from (8.44) that

$$\begin{aligned}\varphi''(u) &= \pi\sqrt{\pi}\left(\phi'(u|1,0.5,\eta) + \phi'(u|0,1.5,\eta')\right) \\ &= \pi^2\sqrt{\pi}\left(\phi(u|0,1.5,\eta) + \phi(u|2,0.5,\eta)\right. \\ &\quad \left.+4\phi(u|1,1.5,\eta') + \phi(u|0,2.5,\eta'')\right), \qquad (8.46)\end{aligned}$$

$$\begin{aligned}\varphi'''(u) &= \pi^2\sqrt{\pi}\,(\phi'(u|0,1.5,\eta)+\phi'(u|2,0.5,\eta)\\ &\quad +4\phi'(u|1,1.5,\eta')+\phi'(u|0,2.5,\eta''))\\ &= \pi^3\sqrt{\pi}\,(5\phi(u|1,1.5,\eta)+\phi(u|3,0.5,\eta)\\ &\quad +5\phi(u|0,2.5,\eta')+13\phi(u|2,1.5,\eta')\\ &\quad +9\phi(u|1,2.5,\eta'')+\phi(u|0,3.5,\eta'''))\,, \end{aligned} \tag{8.47}$$

and

$$\begin{aligned}\varphi''''(u) &= \pi^3\sqrt{\pi}\,(5\phi'(u|1,1.5,\eta)+\phi'(u|3,0.5,\eta)\\ &\quad +5\phi'(u|0,2.5,\eta')+13\phi'(u|2,1.5,\eta')\\ &\quad +9\phi'(u|1,2.5,\eta'')+\phi'(u|0,3.5,\eta'''))\\ &= \pi^4\sqrt{\pi}\,(5\phi(u|0,2.5,\eta)+18\phi(u|2,1.5,\eta)+\phi(u|4,0.5,\eta))\\ &\quad +56\phi(u|1,2.5,\eta')+40\phi(u|3,1.5,\eta')\\ &\quad +58\phi(u|2,2.5,\eta'')+14\phi(u|0,3.5,\eta'')\\ &\quad +16\phi(u|1,3.5,\eta''')+\phi(u|0,4.5,\eta''''))\,. \end{aligned} \tag{8.48}$$

Next, observe from (8.42) that

$$\lim_{u\downarrow 0}\phi(u|k,c,\psi)=\lim_{x\to-\infty}x^{k+2c}\psi\,(x)\,,\ \lim_{u\uparrow 1}\phi(u|k,c,\psi)=\lim_{x\to+\infty}x^{k+2c}\psi\,(x)\,, \tag{8.49}$$

hence by (8.43), (8.45), (8.46) and (8.47),

$$\begin{aligned}\varphi(0) &= \sqrt{\pi}\lim_{x\to-\infty}x\eta\,(x)\,,\\ \varphi(1) &= \sqrt{\pi}\lim_{x\to+\infty}x\eta\,(x)\,,\end{aligned}$$

$$\begin{aligned}\varphi'(0) &= \pi\sqrt{\pi}\left(\lim_{x\to-\infty}x^2\eta\,(x)+\lim_{x\to-\infty}x^3\eta'\,(x)\right),\\ \varphi'(1) &= \pi\sqrt{\pi}\left(\lim_{x\to+\infty}x^2\eta\,(x)+\lim_{x\to+\infty}x^3\eta'\,(x)\right),\end{aligned}$$

$$\begin{aligned}\varphi''(0) &= \pi^2\sqrt{\pi}\left(\lim_{x\to-\infty}x^3\eta\,(x)+4\lim_{x\to-\infty}x^4\eta'\,(x)+\lim_{x\to-\infty}x^5\eta''\,(x)\right),\\ \varphi''(1) &= \pi^2\sqrt{\pi}\left(\lim_{x\to+\infty}x^3\eta\,(x)+4\lim_{x\to+\infty}x^4\eta'\,(x)+\lim_{x\to+\infty}x^5\eta''\,(x)\right),\end{aligned}$$

and

$$\begin{aligned}\varphi'''(0) &= \pi^3\sqrt{\pi}\left(6\lim_{x\to-\infty}x^4\eta(x)+17\lim_{x\to-\infty}x^5\eta'(x)\right.\\&\left.\qquad+9\lim_{x\to-\infty}x^6\eta''(x)+\lim_{x\to-\infty}x^7\eta'''(x)\right),\\ \varphi'''(1) &= \pi^3\sqrt{\pi}\left(6\lim_{x\to+\infty}x^4\eta(x)+17\lim_{x\to+\infty}x^5\eta'(x)\right.\\&\left.\qquad+9\lim_{x\to+\infty}x^6\eta''(x)+\lim_{x\to+\infty}x^7\eta'''(x)\right).\end{aligned}$$

Thus,

$$\begin{array}{l}\varphi'''(0)=\varphi'''(1)=\varphi'(0)=\varphi'(1)=0\\ \text{and } \varphi''(0)=\varphi''(1)=\varphi(0)=\varphi(1)=0\end{array}\tag{8.50}$$

if

$$\lim_{|x|\to\infty}x^4\eta(x)=0,\ \lim_{|x|\to\infty}x^5\eta'(x)=0,\ \lim_{|x|\to\infty}x^6\eta''(x)=0,\ \lim_{|x|\to\infty}x^7\eta'''(x)=0.\tag{8.51}$$

Now assume:

**Assumption 8.1.** *Given a density $f(x)$ with support $\mathbb{R}$, the following conditions hold:*
*(a) $f(x)$ is four times continuously differentiable on $\mathbb{R}$;*
*(b) $\int_{-\infty}^{\infty}|x|^3\sqrt{f(x)}\mathrm{d}x<\infty$;*
*(c) Denoting $\eta(x)=\sqrt{f(x)}$, the set $\{x\in\mathbb{R}:\eta'(x)=0$ or $\eta''(x)=0$ or $\eta'''(x)=0\}$ is finite.*

Assumption 8.1 holds for most densities considered in parametric econometric models, in particular the densities of the normal and logistic distributions.

Now $\lim_{|x|\to\infty}x^4\eta(x)=0$ follows from condition (b) and Lemma 8.1. Due to condition (c) and the fact that $\lim_{|x|\to\infty}\eta(x)=0$ there exists an $a>0$ such that $\eta'(x)\le 0$ for all $x\ge a$ and $\eta'(x)\ge 0$ for all $x\le -a$. Then for $b>a$, using integration by parts,

$$\int_a^b x^4\eta'(x)\mathrm{d}x=b^4\eta(b)-a^4\eta(a)-4\int_a^b x^3\eta(x)\mathrm{d}x.$$

Letting $b \to \infty$ it follows from $\lim_{|x|\to\infty} x^4\eta(x) = 0$ and condition (b) that

$$\int_a^\infty x^4|\eta'(x)|\mathrm{d}x = a^4\eta(a) + 4\int_a^b x^3\eta(x)\mathrm{d}x < \infty,$$

and similarly

$$\int_{-\infty}^{-a} x^4|\eta'(x)|\mathrm{d}x = a^4\eta(-a) + 4\int_{-\infty}^{-a} |x|^3\eta(x)\mathrm{d}x < \infty.$$

Hence, $\int_{-\infty}^{\infty} x^4|\eta'(x)|\mathrm{d}x < \infty$, which by Lemma 8.1 implies that $\lim_{|x|\to\infty} x^5 \eta'(x) = 0$. Next, it follows from part (c) of Assumption 8.1 that for some $a > 0$, $\eta''(x) \neq 0$ for all $x \in [a,\infty)$, and since $\eta'(x) \uparrow 0$ as $x \to \infty$, $\eta'(x) \downarrow 0$ as $x \to -\infty$, we have $\eta''(x) > 0$ for all $x \in [a,\infty)$ and $\eta''(x) < 0$ for all $x \in (-\infty,-a]$. Then

$$\begin{aligned}\int_a^\infty x^5\eta''(x)\mathrm{d}x &= \lim_{b\to\infty} b^5\eta'(b) - a^5\eta'(a) - 5\int_a^\infty x^4\eta'(x)\mathrm{d}x \\ &= -a^5\eta'(a) + 5\int_a^\infty x^4|\eta'(x)|\mathrm{d}x < \infty,\end{aligned}$$

where the second equality follows from Lemma 8.1, and

$$\int_{-\infty}^{-a} x^5\eta''(x)\mathrm{d}x = a^5\eta'(-a) + 5\int_a^\infty x^4|\eta'(x)|\mathrm{d}x < \infty,$$

hence $\int_{-\infty}^{\infty} |x|^5|\eta''(x)|\mathrm{d}x < \infty$, which by Lemma 8.1 implies $\lim_{|x|\to\infty} x^6\eta''(x) = 0$. Similarly, it follows that $\int_{-\infty}^{\infty} x^6|\eta'''(x)|\mathrm{d}x < \infty$, which implies $\lim_{|x|\to\infty} x^7 \eta'''(x) = 0$. Thus,

**Lemma 8.3**. *Assumption 8.1 implies that* $\int_{-\infty}^{\infty} x^4|\eta'(x)|\mathrm{d}x < \infty$, $\int_{-\infty}^{\infty} |x|^5 |\eta''(x)|\mathrm{d}x < \infty$ *and* $\int_{-\infty}^{\infty} x^6|\eta'''(x)|\mathrm{d}x < \infty$, *which by Lemma 8.1 imply that the limits* (8.51) *hold, hence the tail conditions* (8.50) *hold.*

Finally, we have to set forth conditions such that $\varphi''''(u) \in L^2(0,1)$, which is the case if $\int_0^1 \varphi''''(u)^2\mathrm{d}u < \infty$. A sufficient condition for the latter is that

$$\max\left(\lim_{u\downarrow 0}|\varphi''''(u)|, \lim_{u\uparrow 1}|\varphi''''(u)|\right) < \infty, \tag{8.52}$$

because then by part (a) of Assumption 8.1, $\varphi''''(u)$ is uniformly continuous and thus bounded on $[0,1]$.

Now observe from (8.49) and (8.48) that

$$\begin{aligned}
\varphi''''(0) &= \lim_{u\downarrow 0}\varphi''''(u) \\
&= \pi^4\sqrt{\pi}\left(24\lim_{x\to-\infty}x^5\eta(x)+96\lim_{x\to-\infty}x^6\eta'(x)\right. \\
&\quad \left.+72\lim_{x\to-\infty}x^7\eta''(x)+16\lim_{x\to-\infty}x^8\eta'''(x)+\lim_{x\to-\infty}x^9\eta''''(x)\right) \\
\varphi''''(1) &= \lim_{u\uparrow 1}\varphi''''(u) \\
&= \pi^4\sqrt{\pi}\lim_{x\to+\infty}\left(24\lim_{x\to+\infty}x^5\eta(x)+96\lim_{x\to+\infty}x^6\eta'(x)\right. \\
&\quad \left.+72\lim_{x\to+\infty}x^7\eta''(x)+16\lim_{x\to+\infty}x^8\eta'''(x)+\lim_{x\to+\infty}x^9\eta''''(x)\right),
\end{aligned}$$

hence, (8.52) holds if

$$\begin{array}{ll}
\lim_{|x|\to\infty}|x|^5\eta(x)<\infty, & \lim_{|x|\to\infty}x^6|\eta'(x)|<\infty, \\
\lim_{|x|\to\infty}|x^7\eta''(x)|<\infty, & \lim_{|x|\to\infty}x^8|\eta'''(x)|<\infty, \\
\lim_{|x|\to\infty}|x^9\eta''''(x)|<\infty. &
\end{array} \tag{8.53}$$

However, it is difficult, if not impossible, to derive more primitive general conditions for (8.53). On the other hand, suppose that:

**Assumption 8.2.** *In addition to the conditions in Assumption 8.1 the following conditions hold:*
(1) $\int_{-\infty}^{\infty}x^4\sqrt{f(x)}\mathrm{d}x<\infty$;
(2) *The set* $\{x\in\mathbb{R}:\eta''''(x)=0\}$ *is finite, where* $\eta(x)=\sqrt{f(x)}$.

Then it follows similar to Lemma 8.3 that

$$\begin{array}{ll}
\lim_{|x|\to\infty}|x|^5\eta(x)=0, & \lim_{|x|\to\infty}x^6|\eta'(x)|=0, \\
\lim_{|x|\to\infty}|x^7\eta''(x)|=0, & \lim_{|x|\to\infty}x^8|\eta'''(x)|=0, \\
\lim_{|x|\to\infty}|x^9\eta''''(x)|=0, &
\end{array}$$

hence

$$\varphi''''(0)=\varphi''''(1)=0. \tag{8.54}$$

Part (2) of Assumption 8.2 is not a big deal, but admittedly, part (1) is too strong a condition for $\int_0^1 \varphi''''(u)^2 \mathrm{d}u < \infty$. Nevertheless, I will adopt Assumption 8.2, as it holds for most densities considered in parametric econometric models.

Summarizing, the following results hold.

**Theorem 8.6**. *Under Assumptions 8.1 and 8.2 the function* $\varphi(u) = \sqrt{h(u)}$ *defined in* (8.41) *satisfies the tail conditions* (8.50) *and* (8.54) *Consequently, the results in Theorem 8.5 carry over to the density* $h(u)$ *defined by* (8.9).

**Remark**. It can be shown by (8.4), (8.55) and the well-known sine-cosine formulas that for $m \in \mathbb{N}$,

$$\begin{aligned}
\delta_{2m} &= \sqrt{2}(-1)^m \frac{\int_{-\infty}^{\infty} \cos(2m\arctan(x)) \frac{\sqrt{f(x)}}{\sqrt{1+x^2}} \mathrm{d}x}{\int_{-\infty}^{\infty} \frac{\sqrt{f(x)}}{\sqrt{1+x^2}} \mathrm{d}x}, \\
\delta_{2m-1} &= \sqrt{2}(-1)^m \frac{\int_{-\infty}^{\infty} \sin((2m-1)\arctan(x)) \frac{\sqrt{f(x)}}{\sqrt{1+x^2}} \mathrm{d}x}{\int_{-\infty}^{\infty} \frac{\sqrt{f(x)}}{\sqrt{1+x^2}} \mathrm{d}x},
\end{aligned}$$

Hence, if $f(x)$ is symmetric around zero then $\delta_{2m-1} = 0$.

## 8.5 A numerical example

Recall that Assumptions 8.1 and 8.2 hold for the normal density. Therefore, I will base a numerical experiment on the case where $f(x)$ is the density of the $N(\mu, 1)$ distribution, where $\mu = 0.25$. The reason for choosing $\mu \neq 0$ is to avoid $\delta_{2m-1} \equiv 0$, as explained in the last remark.

It is too difficult, if not impossible, to compute the $\delta_m$'s exactly, but they can be approximated arbitrarily close by

$$\delta_m(K) = \sqrt{2} \frac{\sum_{i=0}^{K} \cos(m\pi.i/K)\, \varphi(i/K)}{\sum_{i=0}^{K} \varphi(i/K)}, \ m \in \mathbb{N},$$

for some large $K \in \mathbb{N}$, where $\varphi(u) = \sqrt{h(u)}$ defined in (8.41). In particular, it is not hard to verify that $\lim_{K\to\infty} \delta_m(K) = \delta_m$. Therefore, the values of $\delta_m$

have been approximated by $\delta_m(K)$ for $K = 10,000$ and $m = 1, 2, ..., 50$.[2] It appears that indeed the $\delta_m$'s for odd $m$ are all nonzero.

In Figures 8.1 and 8.2 the density function $h(u) = \varphi(u)^2$ with $\varphi(u)$ given in (8.41) is compared with its SNP version

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

for truncation orders $n = 10$ and 20.

*Insert*
Figure 8.1: $h(u)$ compared with $h_{10}(u)$
*about here.*

*Insert*
Figure 8.2: $h(u)$ compared with $h_{20}(u)$
*about here.*

As we see the SNP approximation $h_n(u)$ fits $h(u)$ quite well. In particular, $h_n(u)$ and $h(u)$ are almost indistinguishable for $n = 20$.

In Figures 8.3 and 8.4 the first derivative $h_n'(u)$ of $h_n(u)$ is compared with the first derivative $h'(u)$ of $h(u)$, again for $n = 10$ and 20, respectively.

*Insert*
Figure 8.3: $h'(u)$compared with $h_{10}'(u)$
*about here.*

*Insert*
Figure 8.4: $h'(u)$ compared with $h_{20}'(u)$
*about here.*

The fit of $h_n'(u)$ for $n = 10$ is rather poor, but improves substantially for $n = 20$, as expected from Theorems 8.5 and 8.6.

Figures 8.5 and 8.6 compare $h_n''(u)$ with $h''(u)$ for $n = 10$ and 20, respectively.

---

[2] Which took only a few seconds on my Lenovo ThinkPad.

*Insert*
Figure 8.5: $h''(u)$ compared with $h''_{10}(u)$
*about here.*

*Insert*
Figure 8.6: $h''(u)$ compared with $h''_{20}(u)$
*about here.*

The fit of $h''_{10}(u)$ is really bad, but at least $h''_{20}(u)$ captures the extrema of $h''(u)$ quite well, although $h''_{20}(u)$ still wiggles too much in between these extrema.

To check whether the latter phenomenon is a permanent feature or is due to too low a truncation order $n$, I have also done the comparison of $h''_n(u)$ with $h''(u)$ for $n = 50$, in Figure 8.7:

*Insert*
Figure 8.7: $h''(u)$ compared with $h''_{50}(u)$
*about here.*

Now the fit is almost perfect!

In conclusion, the approach in this chapter substantially improves the fit of the SNP density $h_n(u)$ based on the cosine series, and its derivatives. In particular, this numerical experiment demonstrates the differences of the rates of uniform convergence of $h_n(u)$, $h'_n(u)$ and $h''_n(u)$ to $h(u)$, $h'(u)$ and $h''(u)$, respectively, as derived in Theorems 8.5.and 8.6.

## 8.6 Densities on $\mathbb{R}$

As have been shown, every density $f_0(x)$ with support $\mathbb{R}$ satisfying Assumptions 8.1 and 8.2 can be written as

$$f_0(x) = f(x|\boldsymbol{\delta}^0) = \frac{h\left(0.5 + \pi^{-1}\arctan(x)|\boldsymbol{\delta}^0\right)}{\pi\left(1 + x^2\right)}, \tag{8.55}$$

where $h_0(u) = h(u|\boldsymbol{\delta}^0)$, with $h(u|\boldsymbol{\delta})$ defined by (8.2), satisfies the conditions of Theorem 8.6, so that

$$f_0(x) = f(x|\boldsymbol{\delta}^0) \equiv \frac{\left(1+\sum_{m=1}^{\infty}\delta_{0,m}\sqrt{2}\cos\left(m\pi(0.5+\pi^{-1}\arctan(x))\right)\right)^2}{\pi\left(1+x^2\right)\left(1+\sum_{k=1}^{\infty}\delta_{0,k}^2\right)} \tag{8.56}$$

exactly and uniformly on $\mathbb{R}$. Moreover, it follows from (8.40) that $\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty}$ satisfies $\sum_{m=1}^{\infty} m^3|\delta_{0,m}| < \infty$.

It follows now similar to Theorems 8.5 and 8.6 that the following theorem holds.

**Theorem 8.7**. *Every density $f_0(x)$ with support $\mathbb{R}$ satisfying Assumptions 8.1 and 8.2 has the exact and uniform cosine series representations $f_0(x) = f(x|\boldsymbol{\delta}^0)$ in (8.55), where*

$$\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty} \in \Delta_3 \stackrel{\text{def.}}{=} \left\{\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} m^3|\delta_m| < \infty\right\}. \tag{8.57}$$

*Moreover, with $\pi_n$ the truncation operator, we have*

$$\begin{aligned}
\sup_{x\in\mathbb{R}}\left|f(x|\boldsymbol{\delta}^0) - f(x|\pi_n\boldsymbol{\delta}^0)\right| &= o(n^{-7/2}), \qquad (8.58)\\
\sup_{x\in\mathbb{R}}\left|f'(x|\boldsymbol{\delta}^0) - f'(x|\pi_n\boldsymbol{\delta}^0)\right| &= o(n^{-5/2}),\\
\sup_{x\in\mathbb{R}}\left|f''(x|\boldsymbol{\delta}^0) - f''(x|\pi_n\boldsymbol{\delta}^0)\right| &= o(n^{-3/2}),\\
\sup_{x\in\mathbb{R}}\left|f'''(x|\boldsymbol{\delta}^0) - f'''(x|\pi_n\boldsymbol{\delta}^0)\right| &= o(n^{-1/2}),
\end{aligned}$$

*where $f_0'(x) = f'(x|\boldsymbol{\delta}^0)$, $f_0''(x) = f''(x|\boldsymbol{\delta}^0)$, and $f_0'''(x) = f'''(x|\boldsymbol{\delta}^0)$, respectively. Furthermore, denoting the c.d.f. of $f(x|\boldsymbol{\delta}^0)$ by $F(x|\boldsymbol{\delta}^0)$, it follows from* (8.58) *that*

$$\sup_{x\in\mathbb{R}}\left|F(x|\boldsymbol{\delta}^0) - F(x|\pi_n\boldsymbol{\delta}^0)\right| = o(n^{-7/2}).$$

# 8.7 Proofs

## 8.7.1 Lemma 8.1

There exists an $x_0 > 0$ such that either $\psi(x) > 0$ or $\psi(x) < 0$ for all $x > x_0$. Without loss of generality we may assume that the former case applies. Thus, $\psi(x) > 0$ for all $x > x_0$. Now suppose that for some $M \in (0, \infty)$, $\lim_{y\to\infty} \sup_{x\geq y} x\psi(x) \geq M$. Then there exists a $y_0 > x_0$ such that for all $x \geq y_0$, $x\psi(x) > M/2$, which implies that $\int_{y_0}^{\infty} \psi(x)\mathrm{d}x \geq (M/2) \int_{y_0}^{\infty} x^{-1}\mathrm{d}x = \infty$. However, the latter contradicts $\int_{-\infty}^{\infty} |\psi(x)|\mathrm{d}x < \infty$. Thus, $\lim_{y\to\infty} \sup_{x\geq y} x\psi(x) = 0$, which implies that $\lim_{x\to\infty} x\psi(x) = 0$. By a similar argument it can be shown that $\lim_{x\to-\infty} x\psi(x) = 0$.

### 8.7.2 Lemma 8.2

For $K \in \mathbb{N}$ and $c > 1/2$, we have by Lyapunov's inequality,

$$\begin{aligned}\frac{\sum_{m=n+1}^{n+K} m^{-c}|\alpha_m|}{\sum_{k=n+1}^{k+K} k^{-2c}} &= \sum_{m=n+1}^{n+K} \left(\frac{m^{-2c}}{\sum_{k=n+1}^{k+K} k^{-2c}}\right) m^{c}|\alpha_m| \\ &\leq \sqrt{\sum_{m=n+1}^{n+K} \left(\frac{m^{-2c}}{\sum_{k=n+1}^{k+K} k^{-2c}}\right) m^{2c}\alpha_m^2} \\ &= \frac{\sqrt{\sum_{m=n+1}^{n+K} \alpha_m^2}}{\sqrt{\sum_{k=n+1}^{k+K} k^{-2c}}}. \end{aligned} \tag{8.59}$$

Letting $K \to \infty$ it follows that

$$\sum_{m=n+1}^{\infty} m^{-c}|\alpha_m| \leq \sqrt{\sum_{m=n+1}^{\infty} \alpha_m^2}\sqrt{\sum_{k=n+1}^{\infty} k^{-2c}}. \tag{8.60}$$

Since $\sum_{k=n+1}^{\infty} k^{-2c} \leq \sum_{k=n+1}^{\infty} \int_{k-1}^{k} x^{-2c}\mathrm{d}x = \int_{n}^{\infty} x^{-2c}\mathrm{d}x = \frac{1}{2c-1} n^{1-2c}$ and $\sum_{m=n+1}^{\infty} \alpha_m^2 = o(1)$, Lemma 8.2 follows.

# Part IV

# Semi-nonparametric modeling and inference

# Chapter 9

# The SNP Tobit model: Consistency

## 9.1 Introduction

In this chapter I will show how to model the well-known Tobit model semi-nonparametrically, derive mild conditions under which the SNP Tobit model is identified, and show that this model can be estimated consistently by sieve maximum likelihood estimation, using the results in Bierens (2014).

A Google Scholar search of "sieve estimation of the semi-nonparametric Tobit model" produced no relevant references, and a Google Scholar search of "semi-nonparametric Tobit model" mainly produced references to "semi-parametric estimation" in general and "semiparametric estimation of Tobit models" in a few cases. Since in the statistical literature the Tobit model is also referred to as a censored regression model, I also tried a Google Scholar search of "semi-nonparametric censored regression models", but that did not result in additional relevant references.

As quoted from Powell (1994), "Semiparametric modeling is, as its name suggests, a hybrid of the parametric and nonparametric approaches to construction, fitting, and validation of statistical models." See also Powell (2008) for a more recent review of semiparametric modeling and inference. Semiparametric models and semi-nonparametric models have in common that parts of these models are left unspecified. These parts usually take the form of unknown functions, for example density and/or distribution functions. In semiparametric modeling these unknown functions are approximated either by nonparametric estimators, or their use is avoided all together, except for certain aspects of these functions. In that respect a regression model is semiparametric because it hinges mainly on the condition that the conditional expectation of the error term given the vector of stochastic regressors is zero with probability 1. The same applies to the semiparametric estimation approaches of Powell (1984, 1986). In the context of the Tobit model, Powell (1984) proposed a conditional least absolute deviations (CLAD) estimator under the condition that the error term in the latent variable model involved has zero conditional median, and Powell (1986) showed that if the distribution of the latter error term is symmetric around zero then the parameters of the Tobit model can be estimated via a least squares approach. In both cases there is no need to specify the error distribution further. More recently, Khan and Powell (2001) improved on Powell's (1984) approach by using nonparametric kernel estimation, and Lewbel and Linton (2002) showed how to estimate a fully nonparametric version of the Tobit model.

In semi-nonparametric models the unknown functions involved are approximated by series expansions, so that these models become fully parametric, albeit with infinite dimensional parameters. The paper by Duncan (1986) falls in the latter category, despite calling his approach "semi-parametric". He proposed to approximate the error density and distribution function of the Tobit model by a spline with mesh size approaching zero with the sample size, and estimate the parameters of the latent variable model together with the knots of the spline approximation by sieve maximum likelihood. However, only consistency is proved. As to asymptotic normality of the sieve estimators of the latent variable Tobit model parameters, Duncan (1986) stated that "A complete asymptotic distribution theory is unavailable." Ding and Nan (2011) consider an SNP censored regression model for duration data, which is somewhat similar to the Tobit model. These authors also use spline sieve estimation, and they also derive asymptotic normality results, but under high-level conditions.

My aim in this and the next two chapters is to demonstrate the applicability of my sieve ML approach under low-level conditions in Bierens (2014) to the SNP Tobit model.

## 9.2 The original parametric Tobit model

The original parametric Tobit[1] model proposed by Tobin (1958) assumes that the observed dependent variable $Y$ is censored from below by zero, i.e.,

$$Y = \max\left(Y^*, 0\right), \tag{9.1}$$

where $Y^*$ is a latent variable generated by the classical linear regression model

$$Y^* = \theta_0' X + V, \tag{9.2}$$

with $X$ a vector of regressors, possibly including 1 for the intercept, and $\theta_0$ the corresponding vector of parameters. The model error $V$ is assumed to be $N(0, \sigma_0^2)$ distributed, conditional on $X$.

Denoting the density of the $N(0,1)$ distribution by $\phi(z)$ with corresponding c.d.f. $\Phi(z)$, the conditional c.d.f. of $Y$ given $Y > 0$ and $X$ is

$$\begin{aligned}
\Psi(y|Y &> 0, X, \theta_0, \sigma_0) = \Pr(Y \leq y|Y > 0, X) \\
&= \frac{\Pr(0 < Y^* \leq y|X)}{\Pr(Y^* > 0|X)} = \frac{\Pr(-\theta_0' X < V \leq y - \theta_0' X|X)}{\Pr(V > -\theta_0' X|X)} \\
&= \frac{\Phi\left((y - \theta_0' X)/\sigma_0\right) - \Phi\left(-\theta_0' X/\sigma_0\right)}{1 - \Phi\left(-\theta_0' X/\sigma_0\right)} \\
&= \frac{\Phi\left((y - \theta_0' X)/\sigma_0\right) - \Phi\left(-\theta_0' X/\sigma_0\right)}{\Phi\left(\theta_0' X/\sigma_0\right)}, \ y > 0,
\end{aligned}$$

where the latter equality follows from the symmetry of the standard normal distribution. The corresponding conditional density is

$$\psi(y|Y > 0, X, \theta_0, \sigma_0) = \frac{\phi\left((y - \theta_0' X)/\sigma_0\right)}{\sigma_0 \Phi\left(\theta_0' X/\sigma_0\right)}, \ y > 0.$$

Moreover,

$$\begin{aligned}
\Pr\left[Y = 0|X\right] &= \Pr[V \leq -\theta_0' X|X] = \Pr[V/\sigma_0 \leq -\theta_0' X/\sigma_0|X] \\
&= \Phi\left(-\theta_0' X/\sigma_0\right).
\end{aligned}$$

[1] The model is called **Tobit** because it was first proposed by **Tob***in* (1958), and involves aspects of Pr*o***bit** analysis.

### 9.2.1 Truncation bias

Now the conditional expectation of $Y$ given $X$ and $Y > 0$ is

$$\begin{aligned}
E\left[Y|X, Y>0\right] &= \int_0^\infty y\psi(y|Y>0, X, \theta_0, \sigma_0)\mathrm{d}y \\
&= \frac{1}{\sigma_0\Phi\left(\theta_0'X/\sigma_0\right)}\int_0^\infty y\phi\left((y-\theta_0'X)/\sigma_0\right)\mathrm{d}y \\
&= \frac{1}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_0^\infty y\phi\left((y-\theta_0'X)/\sigma_0\right)\mathrm{d}(y/\sigma_0) \\
&= \frac{\sigma_0}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_0^\infty z\phi\left(z-\theta_0'X/\sigma_0\right)\mathrm{d}z \\
&= \frac{\sigma_0}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_0^\infty \left(z-\theta_0'X/\sigma_0\right)\phi\left(z-\theta_0'X/\sigma_0\right)\mathrm{d}z \\
&\quad +\frac{\theta_0'X}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_0^\infty \phi\left(z-\theta_0'X/\sigma_0\right)\mathrm{d}z \\
&= \frac{\theta_0'X}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_{-\theta_0'X/\sigma_0}^\infty \phi\left(z\right)\mathrm{d}z \\
&\quad +\frac{\sigma_0}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_{-\theta_0'X/\sigma_0}^\infty z\phi\left(z\right)\mathrm{d}z \\
&= \theta_0'X + \frac{\sigma_0}{\Phi\left(\theta_0'X/\sigma_0\right)}\int_{-\theta_0'X/\sigma_0}^\infty z\phi\left(z\right)\mathrm{d}z.
\end{aligned}$$

Since for the standard normal density $\phi\left(z\right)$, $\phi'\left(z\right) = -z\phi\left(z\right)$, it follows that

$$\begin{aligned}
\int_{-\theta_0'X/\sigma_0}^\infty z\phi\left(z\right)\mathrm{d}z &= -\int_{-\theta_0'X/\sigma_0}^\infty \phi'\left(z\right)\mathrm{d}z = -\left.\phi\left(z\right)\right|_{-\theta_0'X/\sigma_0}^\infty \\
&= \phi\left(-\theta_0'X/\sigma_0\right) = \phi\left(\theta_0'X/\sigma_0\right),
\end{aligned}$$

hence

$$E\left[Y|X, Y>0\right] = \theta_0'X + \frac{\sigma_0.\phi\left(\theta_0'X/\sigma_0\right)}{\Phi\left(\theta_0'X/\sigma_0\right)}.$$

Therefore, if you regress, by OLS, only the positive $Y$'s on the corresponding $X$'s then, due to the latter term, the OLS parameter estimate of $\theta_0$ will be biased. Moreover,

$$\begin{aligned}
E[Y|X] &= E\left[Y|X, Y>0\right].\Pr[Y>0|X] + 0.\Pr[Y\leq 0|X] \\
&= \theta_0'X.\Phi\left(\theta_0'X/\sigma_0\right) + \sigma_0.\phi\left(\theta_0'X/\sigma_0\right)
\end{aligned}$$

so that the linear model is also misspecified if we regress $Y$ on $X$, including the zero $Y$'s.

### 9.2.2 The log-likelihood function

Now the conditional c.d.f. of $Y$ given $X$ only is

$$\begin{aligned}
\Lambda(y|X,\theta_0,\sigma_0) &= \Pr\left[Y \le y|X\right] \\
&= \Pr\left[Y \le y|X, Y>0\right]\Pr\left[Y>0|\ X\right] \\
&\quad + \Pr\left[Y \le y|X, Y=0\right]\Pr\left[Y=0|\ X\right] \\
&= I\left(y>0\right)\Psi(y|Y>0,X,\theta_0,\sigma_0)\Phi\left(\theta_0'X/\sigma_0\right) \\
&\quad + I\left(y=0\right)\left(1-\Phi\left(\theta_0'X/\sigma_0\right)\right),
\end{aligned}$$

where $I(.)$ is the indicator function. Then the function

$$\begin{aligned}
\lambda(y|X,\theta_0,\sigma_0) &= \begin{cases} \partial\Lambda(y|X,\theta_0,\sigma_0)/\partial y & \text{if} \quad y>0 \\ \Lambda(0|X,\theta_0,\sigma_0) & \text{if} \quad y=0 \end{cases} \\
&= I\left(y>0\right)\psi(y|Y>0,X,\theta_0,\sigma_0)\Phi\left(\theta_0'X/\sigma_0\right) \\
&\quad + I\left(y=0\right)\left(1-\Phi\left(\theta_0'X/\sigma_0\right)\right) \\
&= \sigma_0^{-1}I\left(y>0\right)\phi\left((y-\theta_0'X)/\sigma_0\right) \\
&\quad + I\left(y=0\right)\left(1-\Phi\left(\theta_0'X/\sigma_0\right)\right) \\
&= \frac{1}{\sigma_0\sqrt{2\pi}}I\left(y>0\right)\exp\left(-\frac{1}{2}(y-\theta_0'X)^2/\sigma_0^2\right) \\
&\quad + I\left(y=0\right)\left(1-\Phi\left(\theta_0'X/\sigma_0\right)\right),
\end{aligned}$$

is the basis for the likelihood function of the Tobit model. In particular, given a random sample $\{(Y_j,X_j)\}_{j=1}^N$ from the joint distribution of $(Y,X)$, the log-likelihood function of the Tobit model is

$$\begin{aligned}
\mathcal{L}_N(\theta,\sigma) &= \sum_{j=1}^N \ln\left(\lambda(Y_j|X_j,\theta,\sigma)\right) \\
&= \sum_{j=1}^N I(Y_j>0)\left(-\frac{1}{2}(Y_j-\theta'X_j)^2/\sigma^2 - \ln(\sigma)\right) \\
&\quad + \sum_{j=1}^N I(Y_j=0)\ln\left(1-\Phi\left(\theta'X_j/\sigma\right)\right) - \sum_{j=1}^N I(Y_j>0)\ln\left(\sqrt{2\pi}\right).
\end{aligned}$$

See Bierens (2014, section 8.3.4) for a rigorous motivation of the form of $\mathcal{L}_N(\theta, \sigma)$.

However, before maximizing this log-likelihood function it is convenient to reparametrize it by replacing $\sigma$ by $1/\delta$ and $\theta$ by $\sigma\gamma = \gamma/\delta$, so that $\mathcal{L}_N^*(\gamma, \delta) = \mathcal{L}_N(\gamma/\delta, 1/\delta)$ because it has been shown by Olsen (1978) that then the Hessian matrix

$$\frac{\partial^2 \mathcal{L}_N^*(\gamma, \delta)}{\partial(\gamma', \delta)' \partial(\gamma', \delta)}$$

is a.s. negative definite for all values of $\gamma$ and $\delta > 0$. This implies that the log-likelihood $\mathcal{L}_N^*(\gamma, \delta)$ is unimodal. Therefore, the ML estimators of $\gamma_0 = \theta_0/\sigma_0$ and $\delta_0 = 1/\sigma_0$ can be computed very fast by the well-known Newton iteration, and the asymptotic normal distribution of the corresponding estimates of $\theta_0$ and $\sigma_0$ can then be determined using the well-known delta-method. See, for example, Bierens (2014, Theorem 6.25) for the latter.

## 9.3 SNP identification

Instead of assuming zero-mean normality of the error term $V$ in model (9.2), suppose that

**Assumption 9.1**. *Conditional on $X$, the error term $V$ in the latent variable model* (9.2) *has an unknown absolutely continuous distribution with c.d.f. $F_0(v)$, continuous density $f_0(v)$ and support* $\mathbb{R}$, *i.e., $f_0(v) > 0$ for all $v \in \mathbb{R}$.*

Then

$$\Pr[Y = 0|X] = \Pr[V \leq -\theta_0'X|X] = F_0(-\theta_0'X) \tag{9.3}$$

and

$$\Pr[Y \leq y|X, Y > 0] = \frac{F_0(y - \theta_0'X) - F_0(-\theta_0'X)}{1 - F_0(-\theta_0'X)}, \ y > 0, \tag{9.4}$$

with conditional density

$$\partial \Pr[Y \leq y|X, Y > 0]/\partial y = \frac{f_0(y - \theta_0'X)}{1 - F_0(-\theta_0'X)}, \ y > 0. \tag{9.5}$$

In SNP modeling the first question that needs to be asked and answered is: Under what conditions, if any, are the Euclidean parameters ($\theta_0$ in this case) and the non-Euclidean parameter(s) ($F_0$ in this case) identified?

Suppose that there exists a parameter vector $\theta_*$ and an absolutely continuous distribution function $F_*$ with density $f_*$ such that

$$F_0(-\theta_0'X) = F_*(-\theta_*'X) \text{ a.s., and} \tag{9.6}$$

$$\sup_{y>0} |f_0(y - \theta_0'X) - f_*(y - \theta_*'X)| = 0 \text{ a.s.} \tag{9.7}$$

If $X$ contains the constant 1 as its first component, so that $X = (1, X_1')'$ with $X_1$ the vector of stochastic covariates, and with $\theta_0$ and $\theta_*$ partitioned accordingly as $\theta_0 = (\theta_{0,0}, \theta_{0,1}')'$ and $\theta_* = (\theta_{*,0}, \theta_{*,1}')'$, respectively, (9.6) reads $F_0(-\theta_{0,0} - \theta_{0,1}'X_1) = F_*(-\theta_{*,0} - \theta_{*,1}'X_1)$ a.s. Now assume that $\theta_{*,1} = \theta_{0,1}$. Then (9.6) holds with $F_*(v) = F_0(\theta_{*,0} - \theta_{0,0} + v)$. Therefore, either $X$ should not contain a constant, or we need to normalize $F_*$ and $F_0$ somehow, for example by confining $F_*$ and $F_0$ to distribution functions with zero medians, as in Powell (1984, 1986), or zero means.

Let us choose the former option, by assuming that

**Assumption 9.2**. *The vector $X$ of covariates in the SNP Tobit model satisfies $E[X'X] < \infty$ and $\det(\mathrm{Var}(X)) > 0$.*

This assumption excludes the case that one of the components of $X$ is a.s. constant. The condition $E[X'X] < \infty$ guarantees that the variance matrix $\mathrm{Var}(X)$ of $X$ is finite.

In first instance, consider the case that $X \in \mathbb{R}$, so that (9.6) becomes $F_0(-\theta_0.X) = F_*(-\theta_*.X)$ a.s. Clearly, $\theta_0$ and $\theta_*$ must be nonzero and have the same sign, so that with $c = \theta_*/\theta_0 > 0$ and $Z = -\theta_0 X$, (9.6) reads $F_0(Z) = F_*(c.Z)$ a.s., which holds for $F_*(z) = F_0(z/c)$. In the latter case (9.7) reads,

$$\sup_{y>0} \left| f_0(y - \theta_0 X) - c^{-1} f_0(y/c - \theta_0 X) \right| = 0 \text{ a.s.,} \tag{9.8}$$

provided that $X$ (and thus $Z$) has an absolutely continuous distribution with support $\mathbb{R}$. Then by the continuity of $f_0$,

$$f_0(-\theta_0 X) = \lim_{y\downarrow 0} f_0(y - \theta_0 X) = c^{-1} \lim_{y\downarrow 0} f_0(y/c - \theta_0 X) = c^{-1} f_0(-\theta_0 X).$$

Since $f_0(-\theta_0 X) > 0$ a.s. it follows now that $c = 1$ and thus $\theta_* = \theta_0$.

Next, suppose that

**Assumption 9.3**. *$X = (X_1, X_2')'$, where conditional on $X_2$, $X_1$ has an absolutely continuous distribution with support $\mathbb{R}$ and a nonzero coefficient.*

Partitioning $\theta_0$ and $\theta_*$ accordingly as $(\theta_{0,1}, \theta_{0,2}')'$ and $(\theta_{*,1}, \theta_{*,2}')'$ respectively, and denoting $c = \theta_{*,1}/\theta_{0,1}$, the conditions (9.6) and (9.7) read

$$F_0(-\theta_{0,1}X_1 - \theta_{0,2}'X_2) = F_*(-c.\theta_{0,1}X_1 - \theta_{*,2}'X_2) \text{ and } \quad (9.9)$$
$$\sup_{y>0}\left|f_0(y - \theta_{0,1}X_1 - \theta_{0,2}'X_2) - f_*(y - c.\theta_{0,1}X_1 - \theta_{*,2}'X_2)\right| = 0 \quad (9.10)$$

a.s., where $c > 0$. Since conditional on $X_2$, $Z = -\theta_{0,1}X_1$ has an absolutely continuous distribution with support $\mathbb{R}$, (9.9) is equivalent to

$$F_0(z - \theta_{0,2}'X_2) = F_*(c.z - \theta_{*,2}'X_2) \text{ for all } z \in \mathbb{R}, \quad (9.11)$$

and (9.10) is equivalent to

$$f_0(y + z - \theta_{0,2}'X_2) = f_*(y + c.z - \theta_{*,2}'X_2) \text{ for all } z \in \mathbb{R} \text{ and } y > 0. \quad (9.12)$$

Replacing $c.z - \theta_{*,2}'X_2$ in (9.11) by $v$ it follows that

$$F_*(v) = F_0(v/c + (c^{-1}\theta_{*,2} - \theta_{0,2})'X_2) \text{ for all } v \in \mathbb{R}, \quad (9.13)$$

hence

$$f_*(v) = c^{-1}f_0(v/c + (c^{-1}\theta_{*,2} - \theta_{0,2})'X_2) \text{ for all } v \in \mathbb{R},$$

so that (9.12) reads

$$f_0(y + z - \theta_{0,2}'X_2) = c^{-1}f_0(y/c + z - \theta_{0,2}'X_2) \text{ for all } z \in \mathbb{R} \text{ and } y > 0.$$

Substituting $z = \theta_{0,2}'X_2$ in the latter equation yields $f_0(y) = c^{-1}f_0(y/c)$ for all $y > 0$, hence, letting $y \downarrow 0$, it follows that $c = 1$, so that (9.11) becomes

$$F_0(z - \theta_{0,2}'X_2) = F_*(z - \theta_{*,2}'X_2) \text{ for all } z \in \mathbb{R}.$$

Consequently, substituting $z = \theta_{0,2}'X_2$, we have

$$F_0(0) = F_*((\theta_{0,2} - \theta_{*,2})'X_2) = F_*((\theta_0 - \theta_*)'X), \quad (9.14)$$

where the second equality follows from the fact that $c = 1$ implies $\theta_{*,1} = \theta_{0,1}$.

Since the left-hand side of (9.14) does not depend on $X$, neither does the right-hand side. Therefore, $(\theta_0 - \theta_*)'X$ must be a.s. constant, hence $(\theta_0 - \theta_*)'(X - E[X]) = 0$ a.s. and thus

$$(\theta_0 - \theta_*)'\mathrm{Var}(X)(\theta_0 - \theta_*) = 0.$$

By Assumption 9.2 the latter implies $\theta_* = \theta_0$, which in its turn by (9.13) and $c = 1$ implies that $F_* \equiv F_0$.

Summarizing, the following result has been shown.

**Theorem 9.1**. *Under Assumptions 9.1, 9.2 and 9.3 the SNP Tobit model is semi-nonparametrically identified.*

## 9.4 The SNP log-likelihood function

Recall that any absolutely continuous distribution function $F(v)$ on $\mathbb{R}$ and its density $f(v)$ can be written as

$$F(v) = H(G(v)), \; f(v) = h(G(v))g(v),$$

respectively, where $G(v)$ is an a priori chosen absolutely continuous distribution function on $\mathbb{R}$ with continuous density $g(v)$ and support $\mathbb{R}$, and $H(u)$ is an absolutely continuous distribution function on $[0,1]$ with density $h(u)$. Moreover, recall from Theorem 7.4 that $h(u)$ has the SNP representation

$$\begin{aligned} h(u) \quad &= \quad h(u|\boldsymbol{\delta}) \stackrel{\text{def.}}{=} \frac{\left(1 + \sum_{m=1}^{\infty} \delta_m \sqrt{2}\cos(m\pi u)\right)^2}{1 + \sum_{k=1}^{\infty} \delta_k^2} \text{ a.e. on } [0,1], \\ & \qquad \text{where } \sum_{k=1}^{\infty} \delta_k^2 < \infty, \end{aligned}$$

for example, where by Theorem 7.5 the sequence $\boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty}$ is unique if $h(u)$ is continuous and positive on $(0,1)$. The latter holds if $f(v)$ is continuous and positive on $\mathbb{R}$. Furthermore, in this case the c.d.f. $H(u|\boldsymbol{\delta}) = \int_0^u h(z|\boldsymbol{\delta})\mathrm{d}z$ has a closed form expression, as the limit of (7.11). Thus, $F$ and $f$ have the SNP representations

$$F(v|\boldsymbol{\delta}) = H(G(v)|\boldsymbol{\delta}), \; f(v|\boldsymbol{\delta}) = h(G(v)|\boldsymbol{\delta})g(v),$$

respectively, where

$$\boldsymbol{\delta} \in \Delta \stackrel{\text{def.}}{=} \left\{ \boldsymbol{\delta} = \{\delta_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \delta_m^2 < \infty \right\}. \tag{9.15}$$

Note that under the conditions of Theorem 9.1 there exists a unique $\boldsymbol{\delta}^0 = \{\delta_{0,m}\}_{m=1}^{\infty} \in \Delta$ such that $F_0(v) = F(v|\boldsymbol{\delta}^0)$ and $f_0(v) = f(v|\boldsymbol{\delta}^0)$ a.e.

At this point I will not yet use the results in chapter 8, which are based on the choice of the standard Cauchy distribution for $G$, but these results will play a key-role in deriving asymptotic normality properties.

Now the conditional c.d.f. of $Y$ given $X$ only is

$$\begin{aligned} \Lambda(y|X,\theta_0,\boldsymbol{\delta}^0) &= \Pr[Y \le y|X] \\ &= \Pr[Y \le y|X, Y > 0] \Pr[Y > 0|\ X] \\ &\quad + \Pr[Y \le y|X, Y = 0] \Pr[Y = 0|\ X] \\ &= I(y > 0) \left( F(y - \theta_0' X|\boldsymbol{\delta}^0) - F(-\theta_0' X|\boldsymbol{\delta}^0) \right) \\ &\quad + I(y = 0) F(-\theta_0' X|\boldsymbol{\delta}^0), \end{aligned}$$

and again the function

$$\begin{aligned} \lambda(y|X,\theta_0,\boldsymbol{\delta}^0) &= \begin{cases} \partial\Lambda(y|X,\theta_0,\boldsymbol{\delta}^0)/\partial y & \text{if } y > 0 \\ \Lambda(0|X,\theta_0,\boldsymbol{\delta}^0) & \text{if } y = 0 \end{cases} \\ &= I(y > 0) f(y - \theta_0' X|\boldsymbol{\delta}^0) + I(y = 0) F(-\theta_0' X|\boldsymbol{\delta}^0) \end{aligned}$$

is the basis for the log-likelihood function, because

$$\begin{aligned} E\left[ \left. \frac{\lambda(Y|X,\theta,\boldsymbol{\delta})}{\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)} \right| X \right] &= E\left[ I(Y > 0) \left. \frac{\lambda(Y|X,\theta,\boldsymbol{\delta})}{\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)} \right| X \right] \\ &\quad + E\left[ I(Y = 0) \left. \frac{\lambda(Y|X,\theta,\boldsymbol{\delta})}{\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)} \right| X \right] \\ &= \int_0^{\infty} f(y - \theta' X|\boldsymbol{\delta}) \mathrm{d}y + F(-\theta' X|\boldsymbol{\delta}) \\ &= \int_{-\infty}^{\infty} f(y|\boldsymbol{\delta}) \mathrm{d}y = 1, \end{aligned}$$

so that by the trivial inequality $\ln(x) < x - 1$ for $x \in (0, 1)$ and $x > 1$,

$$\begin{aligned} &E[\ln(\lambda(Y|X,\theta,\boldsymbol{\delta}))|X] - E\left[\ln(\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)|X\right] \\ &= E\left[ \ln\left( \frac{\lambda(Y|X,\theta,\boldsymbol{\delta})}{\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)} \right) \middle| X \right] \le E\left[ \left. \frac{\lambda(Y|X,\theta,\boldsymbol{\delta})}{\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)} \right| X \right] - 1 = 0. \end{aligned}$$

Thus,

$$E\left[\ln(\lambda(Y|X,\theta,\boldsymbol{\delta}))|X\right] \leq E\left[\ln(\lambda(Y|X,\theta_0,\boldsymbol{\delta}^0)|X\right], \tag{9.16}$$

where by Theorem 9.1 this inequality is strict for $(\theta,\boldsymbol{\delta}) \neq (\theta_0,\boldsymbol{\delta}^0)$.

Given that

**Assumption 9.4**. *We observe a random sample* $\{(Y_j, X_j')'\}_{j=1}^N$ *from the joint distribution of* $(Y, X')'$, *where* $X \in \mathbb{R}^d$,

the log-likelihood function of the SNP Tobit model, divided by $N$, now takes the form

$$\begin{aligned}\widehat{Q}_N(\theta,\boldsymbol{\delta}) &= \frac{1}{N}\sum_{j=1}^N \ln(\lambda(Y_j|X_j,\theta,\boldsymbol{\delta})) \\ &= \frac{1}{N}\sum_{j=1}^N \{I(Y_j>0)\ln\left(f(Y_j-\theta' X_j|\boldsymbol{\delta})\right) \\ &\qquad + I(Y_j=0)\ln\left(F(-\theta' X_j|\boldsymbol{\delta})\right)\},\end{aligned}$$

The usual assumption for nonlinear models is that the Euclidean parameter vector, $\theta$ in our case, is contained in a given compact subset $\Theta$ of $\mathbb{R}^d$ containing $\theta_0$, so that $\boldsymbol{\xi} = (\theta,\boldsymbol{\delta}) = \{\xi_m\}_{m=1}^\infty \in \Theta\times\Delta$, and one can define various metrics on $\Theta\times\Delta$. However, for notational convenience, it will be assumed more generally that

**Assumption 9.5**. $\boldsymbol{\xi} = (\theta,\boldsymbol{\delta}) = \{\xi_m\}_{m=1}^\infty$ *is confined to an infinite-dimensional metric space* $\Xi$ *with metric* $d(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)$.

For example, let $\Xi = \left\{\boldsymbol{\xi} = \{\xi_m\}_{m=1}^\infty : \sum_{m=1}^\infty \xi_m^2 < \infty\right\}$, endowed with the pseudo-Euclidean norm $||\boldsymbol{\xi}|| = \sqrt{\sum_{m=1}^\infty \xi_m^2}$ and associated metric $d(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2) = ||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||$.

In the sequel I will need to narrow down $\Xi$ somehow, but I will postpone this issue until it arises.

## 9.5 Sieve ML estimation

For notational convenience, write

$$\begin{aligned}\varphi(Z_j,\boldsymbol{\xi}) &= \ln(\lambda(Y_j|X_j,\theta,\boldsymbol{\delta})) \\ &= I(Y_j>0)\ln\left(f(Y_j-\theta' X_j|\boldsymbol{\delta})\right) \\ &\qquad + I(Y_j=0)\ln\left(F(-\theta' X_j|\boldsymbol{\delta})\right),\end{aligned} \tag{9.17}$$

where $Z_j=(Y_j,X_j')'$ and $\boldsymbol{\xi}=(\theta,\boldsymbol{\delta})$, so that

$$\widehat{Q}_N(\boldsymbol{\xi})=\widehat{Q}_N(\theta,\boldsymbol{\delta})=\frac{1}{N}\sum_{j=1}^{N}\varphi(Z_j,\boldsymbol{\xi}),$$

and let

$$Q(\boldsymbol{\xi})=E\left[\varphi(Z_j,\boldsymbol{\xi})\right].$$

Note that, with $\boldsymbol{\xi}^0=(\theta_0,\boldsymbol{\delta}^0)$, $Q(\boldsymbol{\xi})=Q(\boldsymbol{\xi})-Q(\boldsymbol{\xi}^0)+Q(\boldsymbol{\xi}^0)\leq Q(\boldsymbol{\xi}^0)$, hence

$$\boldsymbol{\xi}^0=(\theta_0,\boldsymbol{\delta}^0)=\arg\max_{\boldsymbol{\xi}\in\Xi}Q(\boldsymbol{\xi}),$$

which under the conditions of Theorem 9.1 is unique.

## 9.6 Consistency of sieve estimators

The idea of sieve estimation is to construct an increasing sequence of compact subspaces $\Xi_n$ of $\Xi$, called sieve spaces, satisfying $\overline{\cup_{n=1}^{\infty}\Xi_n}=\Xi$, and then estimate $\boldsymbol{\xi}^0$ by

$$\widehat{\boldsymbol{\xi}}_{n_N}=\arg\max_{\boldsymbol{\xi}\in\Xi_{n_N}}\widehat{Q}_N(\boldsymbol{\xi}), \tag{9.18}$$

where $n_N$ is a subsequence of the sample size $N$ such that $\lim_{N\to\infty}n_N/N=0$ and $\lim_{N\to\infty}n_N=\infty$.

The consistency of sieve estimators is well-established in the sieve estimation literature. See Chen (2007) for a recent review. In particular, the sieve consistency proof in Chen (2007, Theorem 3.1) employs the condition that

$$\operatorname*{plim}_{N\to\infty}\sup_{\boldsymbol{\xi}\in\Xi_{n_N}}|\widehat{Q}_N(\boldsymbol{\xi})-Q(\boldsymbol{\xi})|=0. \tag{9.19}$$

The validity of this condition depends on the complexity of the sieve spaces $\Xi_{n_N}$ in terms of covering numbers. See van der Vaart and Wellner (1996) and van der Vaart (1998) for the latter. Moreover, in the log-likelihood case it is possible that $Q(\boldsymbol{\xi}) = -\infty$ for some $\boldsymbol{\xi} \in \Xi$, as is demonstrated by Bierens (2014), and if such a $\boldsymbol{\xi}$ is contained in $\Xi_n$ for large enough $n$ then $\sup_{\boldsymbol{\xi}\in\Xi_{n_N}} |\widehat{Q}_N(\boldsymbol{\xi}) - Q(\boldsymbol{\xi})| \overset{\text{a.s.}}{\to} \infty$. Anyhow, (9.19) is too high-level a condition for my taste, and therefore in Bierens (2014) I have derived the consistency of sieve ML estimators under lower-level and, with one exception, verifiable conditions.

Wald (1949) has proved the consistency of finite-dimensional maximum likelihood estimators without requiring that the expectation of the log-likelihood is everywhere finite. Bahadur (1967) has extended Wald's approach to general compact metric spaces. The more general and transparent Wald consistency proof in van der Vaart (1998, Theorem 5.14) also applies to metric spaces and allows the expectation of the objective function to be minus infinity for some parameter values. Therefore, rather than adopting Theorem 3.1 in Chen (2007), I have, in Bierens (2014), generalized Theorem 5.14 in van der Vaart (1998) to sieve estimators, as follows.

**Assumption 9.6**. *Consider an empirical objective function*

$$\widehat{Q}_N(\boldsymbol{\xi}) = \frac{1}{N}\sum_{j=1}^{N}\varphi(Z_j,\boldsymbol{\xi}),\ \boldsymbol{\xi}\in\Xi,$$

*with population counterpart* $Q(\boldsymbol{\xi}) = E[\varphi(Z,\boldsymbol{\xi})]$, *where*:
(*a*) $Z_1, Z_2, ..., Z_N$ *are independent and identically distributed as* $Z$, *with support contained in an open set* $\mathcal{Z}$ *of a Euclidean space*;
(*b*) $\Xi$ *is an infinite-dimensional metric space with metric* $d(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2)$;
(*c*) *for each* $\boldsymbol{\xi}\in\Xi$, $\varphi(z,\boldsymbol{\xi})$ *is a Borel measurable real function on* $\mathcal{Z}$;
(*d*) $\varphi(Z,\boldsymbol{\xi})$ *is a.s. continuous in* $\boldsymbol{\xi}\in\Xi$;
(*e*) *there exists a non-negative Borel measurable real function* $\overline{\varphi}(z)$ *on* $\mathcal{Z}$ *satisfying* $E[\overline{\varphi}(Z)] < \infty$ *and* $\varphi(Z,\boldsymbol{\xi}) < \overline{\varphi}(Z)$ *a.s. for all* $\boldsymbol{\xi}\in\Xi$;
(*f*) *there exists an element* $\boldsymbol{\xi}^0 \in \Xi$ *such that* $Q(\boldsymbol{\xi}) < Q(\boldsymbol{\xi}^0)$ *for all* $\boldsymbol{\xi} \in \Xi\backslash\{\boldsymbol{\xi}^0\}$, *where* $Q(\boldsymbol{\xi}^0) > -\infty$;
(*g*) *there exists an increasing sequence of compact subspaces* $\Xi_n$ *of* $\Xi$ *such that* $\overline{\cup_{n=1}^{\infty}\Xi_n} = \Xi$.
(*h*) *each sieve space* $\Xi_n$ *is isomorph to a compact subset of a Euclidean space*;
(*i*) *each sieve space* $\Xi_n$ *contains an element* $\boldsymbol{\xi}_n$ *such that* $\lim_{n\to\infty} Q(\boldsymbol{\xi}_n) = Q(\boldsymbol{\xi}^0)$;

*(j) the set* $\Xi_{-\infty} = \{\boldsymbol{\xi} \in \Xi : Q(\boldsymbol{\xi}) = -\infty\}$ *does not contain an open ball, i.e. for each* $\boldsymbol{\xi} \in \Xi_{-\infty}$ *and arbitrary* $\varepsilon > 0$ *there exists a* $\boldsymbol{\xi}_* \in \Xi$ *such that* $d(\boldsymbol{\xi}, \boldsymbol{\xi}_*) < \varepsilon$ *and* $Q(\boldsymbol{\xi}_*) > -\infty$.

Except for condition (j), all the other conditions of Assumption 9.6 hold for the SNP Tobit model. However, it is difficult, if not impossible, to verify condition (j) for the SNP Tobit model, but at least it seems a plausible condition.

Note that due to conditions (d) and (g), $\widehat{\boldsymbol{\xi}}_{n_N} \in \Xi_{n_N}$ a.s., as is not hard to verify.[2] The role of condition (h) is two-fold. First, it makes the computation of $\widehat{\boldsymbol{\xi}}_{n_N}$ feasible, as then the maximization problem (9.18) can be solved in the same way as if $\Xi_{n_N}$ were a compact subset of a Euclidean space itself. Second, it follows from conditions (d) and (h) and Lemma 2 of Jennrich (1969) that suprema of $\varphi(z, \boldsymbol{\xi})$ over subsets of $\Xi_n$ and their limits for $n \to \infty$ are Borel measurable in $z$, and that $d(\widehat{\boldsymbol{\xi}}_{n_N}, \boldsymbol{\xi}^0)$ is a well-defined random variable.

As demonstrated in Bierens (2014), it is possible for SNP log-likelihood models that for some $\boldsymbol{\xi} \in \Xi$, $E[\varphi(Z, \boldsymbol{\xi})] = -\infty$. Condition (j) guarantees that for arbitrary $\varepsilon > 0$,

$$E\left[\sup_{\boldsymbol{\xi}_* \in \Xi,\ d(\boldsymbol{\xi}, \boldsymbol{\xi}_*) < \varepsilon} \varphi(Z, \boldsymbol{\xi}_*)\right] > -\infty$$

if $E[\varphi(Z, \boldsymbol{\xi})] = -\infty$, which of course also applies to the case $E[\varphi(Z, \boldsymbol{\xi})] > -\infty$. Then together with condition (e),

$$E\left[\left|\sup_{\boldsymbol{\xi}_* \in \Xi,\ d(\boldsymbol{\xi}, \boldsymbol{\xi}_*) < \varepsilon} \varphi(Z, \boldsymbol{\xi}_*)\right|\right] < \infty, \tag{9.20}$$

so that by Kolmogorov's strong law of large numbers,

$$\frac{1}{N} \sum_{j=1}^{N} \sup_{\boldsymbol{\xi}_* \in \Xi,\ d(\boldsymbol{\xi}, \boldsymbol{\xi}_*) < \varepsilon} \varphi(Z_j, \boldsymbol{\xi}_*) \stackrel{\text{a.s.}}{\to} E\left[\sup_{\boldsymbol{\xi}_* \in \Xi,\ d(\boldsymbol{\xi}, \boldsymbol{\xi}_*) < \varepsilon} \varphi(Z, \boldsymbol{\xi}_*)\right] \tag{9.21}$$

pointwise in $\boldsymbol{\xi} \in \Xi$.

It follows now similar to Theorem 5.14 in van der Vaart (1998) that the following result holds.

[2] See for example Bierens (2004, Theorem II.6, p.290).

**Theorem 9.2**. *Let the conditions in Assumption 9.6 hold, and let*

$$\widehat{\boldsymbol{\xi}}_{n_N} = \arg\max_{\boldsymbol{\xi}\in\Xi_{n_N}} \widehat{Q}_N(\boldsymbol{\xi})$$

*be the sieve estimator of* $\boldsymbol{\xi}^0 = \arg\max_{\boldsymbol{\xi}\in\Xi} Q(\boldsymbol{\xi})$, *where* $n_N$ *is any subsequence of* $N$ *satisfying* $\lim_{N\to\infty} n_N = \infty$, $\lim_{N\to\infty} n_N/N = 0$. *Then for any* $\varepsilon > 0$ *and any compact subset* $\Xi_c$ *of* $\Xi$,

$$\lim_{N\to\infty} \Pr\left[d\left(\widehat{\boldsymbol{\xi}}_{n_N}, \boldsymbol{\xi}^0\right) \geq \varepsilon \text{ and } \widehat{\boldsymbol{\xi}}_{n_N} \in \Xi_c\right] = 0.$$

*Consequently,* $\text{plim}_{N\to\infty} d\left(\widehat{\boldsymbol{\xi}}_{n_N}, \boldsymbol{\xi}^0\right) = 0$ *if and only if for some infinite-dimensional compact subset* $\Xi_c$ *of* $\Xi$ *with* $\boldsymbol{\xi}^0$ *in its interior,*

$$\lim_{N\to\infty} \Pr\left[\widehat{\boldsymbol{\xi}}_{n_N} \in \Xi_c\right] = 1. \tag{9.22}$$

**Proof.** See the proof of Theorem 4.1 in the online supplement to Bierens (2014), reprinted in Bierens (2017, Ch.10). ■

## 9.7 Compactness

The problem is how to guarantee that (9.22) holds for some infinite-dimensional compact subset $\Xi_c$ of $\Xi$ containing $\boldsymbol{\xi}^0$ in its interior. So the first problem that needs to be addressed is how to construct infinite-dimensional compact spaces.

Recall that a subspace $\Xi_c$ of a metric space $\Xi$ is compact if and only if every open covering of $\Xi_c$ has a finite subcovering. The latter condition is often difficult to verify directly, but there are a few easier to verify conditions for compactness, namely the following.

**Lemma 9.1**. *A subset* $\Xi_c$ *of a metric space* $\Xi$ *is compact if and only if the following two conditions hold.*
*(a)* $\Xi_c$ *is complete, i.e., every Cauchy sequence in* $\Xi_c$ *converges to a limit in* $\Xi_c$;
*(b)* $\Xi_c$ *is totally bounded, i.e. for every* $\varepsilon > 0$, $\Xi_c$ *can be covered by a finite union of spheroids of radius* $\varepsilon$.

**Proof.** See Royden (1968, Proposition 15, p. 164). ■

**Lemma 9.2**. *A subset $\Xi_c$ of a metric space $\Xi$ is compact if and only if it is sequentially compact, i.e., any infinite sequence in $\Xi_c$ has a convergent subsequence with limit contained in $\Xi_c$.*

**Proof.** See Royden (1968, Corollary 14, p.163). ■

The first example of an infinite-dimensional compact space is the following corollary of Lemma A.1 in Bierens (2008), which was originally formulated for spaces of functions on $[0,1]$.

**Lemma 9.3.** *Consider the space $\Xi = \{\boldsymbol{\xi} = \{\xi_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty}\xi_m^2 < \infty\}$ endowed with the pseudo-Euclidean norm $||\boldsymbol{\xi}|| = \sqrt{\sum_{m=1}^{\infty}\xi_m^2}$ and associated metric. Let $\overline{\boldsymbol{\xi}} = \left\{\overline{\xi}_m\right\}_{m=1}^{\infty} \in \Xi$ be a given sequence of positive numbers satisfying $\sum_{m=1}^{\infty}\overline{\xi}_m^2 < \infty$. Then the infinite-dimensional space $\Xi(\overline{\boldsymbol{\xi}}) = \times_{m=1}^{\infty}[-\overline{\xi}_m, \overline{\xi}_m]$ is compact.*

In Bierens (2014, Theorem 4.2) I have shown that if the compact set $\Xi_c$ containing $\boldsymbol{\xi}^0$ in its interior can be chosen so large that

$$E\left[\sup_{\boldsymbol{\xi}\in\Xi\backslash\Xi_c}\varphi(Z,\boldsymbol{\xi})\right] < E[\varphi(Z,\boldsymbol{\xi}^0)]$$

then under the conditions of Theorem 9.2, $\text{plim}_{N\to\infty} d\left(\widehat{\boldsymbol{\xi}}_{n_N}, \boldsymbol{\xi}^0\right) = 0$. Therefore, on the basis of the result of Lemma 9.3, I proposed to choose $\Xi_c$ as

$$\Xi_K = \times_{m=1}^{\infty}[-\overline{\xi}_m K, \overline{\xi}_m K],$$

for some large $K > 1$, where $\{\overline{\xi}_m\}_{m=1}^{\infty}$ is a positive sequence satisfying $\overline{\xi}_m > |\xi_{0,m}|$ for all $m \in \mathbb{N}$ and $\sum_{m=1}^{\infty}\overline{\xi}_m^2 < \infty$. Now if

$$\lim_{K\to\infty} E\left[\sup_{\boldsymbol{\xi}\in\Xi\backslash\Xi_K}\varphi(Z,\boldsymbol{\xi})\right] < E[\varphi(Z,\boldsymbol{\xi}^0)], \tag{9.23}$$

where the limit exists by monotonicity, then for sufficiently large $K$, condition (9.22) holds with $\Xi_c = \Xi_K$, so that with $d(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ the pseudo-Euclidean metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||$, $\text{plim}_{N\to\infty}||\widehat{\boldsymbol{\xi}}_{n_N} - \boldsymbol{\xi}^0|| = 0$.

However, as I have argued in the addendum to Bierens (2014) in Bierens (2017, Ch. 10), this approach is flawed because for each $K > 1$,

$$E\left[\sup_{\boldsymbol{\xi}\in\Xi\backslash\Xi_K}\varphi(Z,\boldsymbol{\xi})\right] \geq E[\varphi(Z,\boldsymbol{\xi}^0)].$$

To see this, let $\boldsymbol{\xi}_n^0 = \{\xi_m\}_{m=1}^\infty$ with $\xi_m = \xi_{0,m}$ for $m \neq n$ and $\xi_n = 2K\overline{\xi}_n$. Note that $\boldsymbol{\xi}_n^0 \in \Xi\backslash\Xi_K$ but $||\boldsymbol{\xi}_n^0 - \boldsymbol{\xi}^0|| = 2K\overline{\xi}_n \to 0$ as $n \to \infty$. Hence, $\sup_{\boldsymbol{\xi}\in\Xi\backslash\Xi_K}\varphi(Z,\boldsymbol{\xi}) \geq \limsup_{n\to\infty}\varphi(Z,\boldsymbol{\xi}_n^0) = \varphi(Z,\boldsymbol{\xi}^0)$ a.s.

Therefore, in the addendum to Bierens (2014) in Bierens (2017, Ch. 10) I proposed an alternative approach based on the following result.

**Lemma 9.4**. *For $\ell > 0.5$, let $\Xi_\ell$ be the space*

$$\Xi_\ell = \left\{\boldsymbol{\xi} = \{\xi_m\}_{m=1}^\infty : \sum_{m=1}^\infty m^\ell|\xi_m| < \infty\right\},$$

*endowed with the norm $||\boldsymbol{\xi}||_\ell = \sum_{m=1}^\infty m^\ell|\xi_m|$ and associated metric. Then for $M \in (0,\infty)$ and $\ell > 0.5$ the space*

$$\Xi_\ell(M) = \left\{\boldsymbol{\xi} = \{\xi_m\}_{m=1}^\infty : \sum_{m=1}^\infty m^\ell|\xi_m| \leq M\right\},$$

*is compact with respect to the metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||_\ell$, as well as with respect to the pseudo-Euclidean metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||$.*

**Remarks.**

1. Lemma 9.4 is in essence a combination of Lemmas 3.1 and 3.2 in Bierens (2017, pp. 632-633), except that the latter lemmas assume $\ell \in \mathbb{N}$. However, all that is needed in Lemma 9.3 and thus in Lemma 9.4 as well is that $\sum_{m=1}^\infty m^{-2\ell} < \infty$, which holds for $\ell > 0.5$.

2. Moreover, note that the proof of Lemma 3.2 in Bierens (2017, pp. 632-633) was incorrect. The proof of Lemma 9.4 below is the corrected version.

Assuming that $\boldsymbol{\xi}^0 \in \Xi_\ell$ for some $\ell > 0.5$, we can always choose $M > ||\boldsymbol{\xi}^0||_\ell$, so that $\boldsymbol{\xi}^0$ is contained in the interior of $\Xi_\ell(M)$. Moreover, since $\Xi_\ell \backslash \Xi_\ell(M)$ is decreasing in $M$ and $\cap_{M>0}(\Xi_\ell \backslash \Xi_\ell(M)) = \emptyset$, it seems plausible that

$$\lim_{M\to\infty} E\left[\sup_{\boldsymbol{\xi}\in\Xi_\ell\backslash\Xi_\ell(M)} \varphi(Z,\boldsymbol{\xi})\right] < E[\varphi(Z,\boldsymbol{\xi}^0)]. \tag{9.24}$$

If so, it follows from Theorem 4.2 in Bierens (2014) that for some large $M$,

$$\lim_{N\to\infty} \Pr\left[\widehat{\boldsymbol{\xi}}_{n_N} \in \Xi_\ell(M)\right] = 1,$$

hence by Theorem 9.2, $\text{plim}_{N\to\infty}||\widehat{\boldsymbol{\xi}}_{n_N} - \boldsymbol{\xi}^0||_\ell = 0$, provided that the sieve spaces involved are defined appropriately, for example as

$$\Xi_{\ell,n} = \left\{\boldsymbol{\xi} = \{\xi_m\}_{m=1}^\infty : \sum_{m=1}^n m^\ell|\xi_m| \le K_n,\ \xi_m = 0 \text{ for } m > n\right\},$$

where $K_n \to \infty$ as $n \to \infty$.

Still, condition (9.24) is difficult to verify, and is of too high a level to my taste. Therefore, in the addendum in Bierens (2017, Ch. 10) I propose to use $\Xi_\ell(M)$ for some $\ell > 0.5$ and $M > ||\boldsymbol{\xi}^0||_\ell$ as the parameter space $\Xi$, with metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||_\ell$, provided that $||\boldsymbol{\xi}^0||_\ell < \infty$. Note that under Assumptions 8.1 and 8.2 the latter condition holds for $\ell = 3$. Then the following result holds.

**Theorem 9.3**. *Suppose that for some $\ell > 0.5$, $||\boldsymbol{\xi}^0||_\ell < \infty$. For this $\ell$, replace in Assumption 9.6 $\Xi$ by $\Xi_\ell(M)$ for some large $M > ||\boldsymbol{\xi}^0||_\ell$, and replace the sieve spaces $\Xi_n$ by*

$$\Xi_{\ell,n}(M) = \left\{\boldsymbol{\xi} = \{\xi_m\}_{m=1}^\infty : \sum_{m=1}^n m^\ell|\xi_m| \le M,\ \xi_m = 0 \text{ for } m > n\right\}.$$

*Then the sieve estimator*

$$\widehat{\boldsymbol{\xi}}_{n_N} = \left(\widehat{\theta}_{n_N}, \widehat{\boldsymbol{\delta}}_{n_N}\right) = \arg\max_{\boldsymbol{\xi}\in\Xi_{\ell,n_N}(M)} \widehat{Q}_N(\boldsymbol{\xi})$$

*of the SNP Tobit model is consistent, i.e.,* $\text{plim}_{N\to\infty}||\widehat{\boldsymbol{\xi}}_{n_N} - \boldsymbol{\xi}^0||_\ell = 0$ *for any subsequence $n_N$ of $N$ satisfying* $\lim_{N\to\infty} n_N = \infty$, $\lim_{N\to\infty} n_N/N = 0$. *Thus,* $\text{plim}_{N\to\infty}\widehat{\theta}_{n_N} = \theta_0$ *and* $\text{plim}_{N\to\infty}\sum_{m=1}^\infty (d+m)^\ell|\widehat{\delta}_{n_N,m} - \delta_{0,m}| = 0$, *hence*

$$\text{plim}_{N\to\infty}||\widehat{\boldsymbol{\delta}}_{n_N} - \boldsymbol{\delta}^0||_\ell = 0. \tag{9.25}$$

**Remark**. Since by Lemma 9.4, $\Xi_\ell(M)$ is also compact with the pseudo-Euclidean metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||$, the result (9.25) also holds as $\text{plim}_{N\to\infty}||\widehat{\boldsymbol{\delta}}_{n_N} - \boldsymbol{\delta}^0|| = 0$.

## 9.8 Asymptotic normality

The parameter vector $\theta_0$ in the SNP Tobit model under review is the main parameter of interest because these parameters measure the effect of the covariates on the latent dependent variable $Y^*$. Therefore, in order to conduct inference on $\theta_0$ it is desirable to establish the asymptotic normality of the sieve estimator $\widehat{\theta}_{n_N}$, i.e.,

$$\sqrt{N}(\widehat{\theta}_{n_N} - \theta_0) \xrightarrow{\text{d}} \mathcal{N}_d(0, \Sigma) \tag{9.26}$$

for some asymptotic variance matrix $\Sigma$.

The error density $f_0(v) = f(v|\boldsymbol{\delta}^0)$ and thus $\boldsymbol{\delta}^0$ as well act as nuisance parameters. Of course, the role these nuisance parameters is crucial because $\theta_0$ depends on $f_0(v)$, and the shape of $f_0(v)$, estimated by $f(v|\widehat{\boldsymbol{\delta}}_{n_N})$, is of course of interest itself.

In the next two chapters I will discuss two approaches for deriving the asymptotic normality result (9.26), with $\Sigma$ the asymptotically efficient variance matrix. The first approach is based on the seminal paper by Shen (1997), and the second approach is based on Bierens (2014).

## 9.9 Proofs

### 9.9.1 Lemma 9.3

By Lemma 9.1 it suffices to prove that $\Xi(\overline{\boldsymbol{\xi}})$ is complete and totally bounded. To prove completeness, let $\boldsymbol{\xi}_n = \times_{m=1}^{\infty}\{\xi_{m,n}\}$ be an arbitrary Cauchy sequence in $\Xi(\overline{\boldsymbol{\xi}})$. Because for each $m \in \mathbb{N}$, $\xi_{m,n}$ is a Cauchy sequence in the compact interval $[-\overline{\xi}_m, \overline{\xi}_m]$, it follows that $\lim_{n\to\infty} \xi_{m,n} = \underline{\xi}_m \in [-\overline{\xi}_m, \overline{\xi}_m]$. See Theorem 2.2. Given an arbitrary $\varepsilon > 0$, we can choose an $M \in \mathbb{N}$ so large that

$$\sum_{m=M+1}^{\infty} (\xi_{m,n} - \underline{\xi}_m)^2 \leq 2 \sum_{m=M+1}^{\infty} \xi_{m,n}^2 + 2 \sum_{m=M+1}^{\infty} \underline{\xi}_m^2 \leq 4 \sum_{m=M+1}^{\infty} \overline{\xi}_m^2 < \varepsilon,$$

whereas obviously, $\lim_{n\to\infty}\sum_{m=1}^{M}(\xi_{m,n}-\underline{\xi}_m)^2 = 0$. Thus, denoting $\underline{\boldsymbol{\xi}} = \{\underline{\xi}_m\}_{m=1}^{\infty}$, it follows trivially that $\lim_{n\to\infty}||\boldsymbol{\xi}_n - \underline{\boldsymbol{\xi}}||^2 = 0$, where $\underline{\boldsymbol{\xi}} \in \Xi(\overline{\boldsymbol{\xi}})$. Hence, $\Xi(\overline{\boldsymbol{\xi}})$ is complete.

To prove total boundedness, observe that for an arbitrary $\varepsilon > 0$,

$$\Xi(\overline{\boldsymbol{\xi}}) \subset \cup_{\boldsymbol{\xi}_*\in\Xi(\overline{\boldsymbol{\xi}})}\{\boldsymbol{\xi}\in\Xi : ||\boldsymbol{\xi}-\boldsymbol{\xi}_*|| < \varepsilon/2\}.$$

Note that for $\boldsymbol{\xi}_* = \{\xi_{*,m}\}_{m=1}^{\infty} \in \Xi(\overline{\boldsymbol{\xi}})$ and $\boldsymbol{\xi} = \{\xi_m\}_{m=1}^{\infty} \in \Xi$,

$$\begin{aligned}\left|\sum_{m=n+1}^{\infty}(\xi_{*,m}-\xi_m)^2 - \sum_{m=n+1}^{\infty}\xi_m^2\right| &= \left|\sum_{m=n+1}^{\infty}\xi_{*,m}(\xi_{*,m}-2\xi_m)\right| \\ &= \left|2\sum_{m=n+1}^{\infty}\xi_{*,m}(\xi_{*,m}-\xi_m) - \sum_{m=n+1}^{\infty}\xi_{*,m}^2\right| \\ &\leq 2\sum_{m=n+1}^{\infty}\overline{\xi}_m|\xi_m-\xi_{*,m}| + \sum_{m=n+1}^{\infty}\overline{\xi}_m^2.\end{aligned}$$

For any integer $M > n$ it follows from the Cauchy-Schwarz inequality that

$$\frac{1}{M-n}\sum_{m=n+1}^{M}\overline{\xi}_m|\xi_m-\xi_{*,m}| \leq \sqrt{\frac{1}{M-n}\sum_{m=n+1}^{M}\overline{\xi}_m^2}\sqrt{\frac{1}{M-n}\sum_{m=n+1}^{M}(\xi_m-\xi_{*,m})^2},$$

hence, multiplying both sides by $M-n$ first and then letting $M\to\infty$ yield

$$\begin{aligned}\sum_{m=n+1}^{\infty}\overline{\xi}_m|\xi_m-\xi_{*,m}| &\leq \sqrt{\sum_{m=n+1}^{\infty}(\xi_m-\xi_{*,m})^2}\sqrt{\sum_{m=n+1}^{\infty}\overline{\xi}_m^2} \\ &\leq \sqrt{\sum_{m=1}^{\infty}(\xi_m-\xi_{*,m})^2}\sqrt{\sum_{m=n+1}^{\infty}\overline{\xi}_m^2} \\ &= ||\boldsymbol{\xi}-\boldsymbol{\xi}_*||\sqrt{\sum_{m=n+1}^{\infty}\overline{\xi}_m^2}.\end{aligned}$$

Thus, given $||\boldsymbol{\xi}-\boldsymbol{\xi}_*|| < \varepsilon/2$, we have

$$\begin{aligned}\left|\sum_{m=n+1}^{\infty}(\xi_{*,m}-\xi_m)^2 - \sum_{m=n+1}^{\infty}\xi_m^2\right| &< \varepsilon\sqrt{\sum_{m=n+1}^{\infty}\overline{\xi}_m^2} + \sum_{m=n+1}^{\infty}\overline{\xi}_m^2 \\ &\to 0 \text{ as } n\to\infty,\end{aligned}$$

hence, $\lim_{n\to\infty} ||\boldsymbol{\xi} - \pi_n \boldsymbol{\xi}_*||^2 = ||\boldsymbol{\xi} - \boldsymbol{\xi}_*||^2$ uniformly in $\boldsymbol{\xi}_* \in \Xi(\overline{\boldsymbol{\xi}})$, where $\pi_n$ is the truncation operator. Consequently, given $||\boldsymbol{\xi} - \boldsymbol{\xi}_*|| < \varepsilon/2$, we can choose $n$ so large that $||\boldsymbol{\xi} - \pi_n \boldsymbol{\xi}_*|| < \varepsilon$, so that for all $\boldsymbol{\xi}_* \in \Xi(\overline{\boldsymbol{\xi}})$,

$$\{\boldsymbol{\xi} \in \Xi : ||\boldsymbol{\xi} - \boldsymbol{\xi}_*|| < \varepsilon/2\} \subset \{\boldsymbol{\xi} \in \Xi : ||\boldsymbol{\xi} - \pi_n \boldsymbol{\xi}_*|| < \varepsilon\}.$$

Therefore,

$$\begin{aligned} \Xi(\overline{\boldsymbol{\xi}}) &\subset \cup_{\boldsymbol{\xi}_* \in \Xi(\overline{\boldsymbol{\xi}})} \{\boldsymbol{\xi} \in \Xi : ||\boldsymbol{\xi} - \boldsymbol{\xi}_*|| < \varepsilon/2\} \\ &\subset \cup_{\boldsymbol{\xi}_* \in \Xi(\overline{\boldsymbol{\xi}})} \{\boldsymbol{\xi} \in \Xi : ||\boldsymbol{\xi} - \pi_n \boldsymbol{\xi}_*|| < \varepsilon\} \\ &= \cup_{\boldsymbol{\xi}_* \in \Xi(\pi_n \overline{\boldsymbol{\xi}})} \{\boldsymbol{\xi} \in \Xi : ||\boldsymbol{\xi} - \boldsymbol{\xi}_*|| < \varepsilon\}, \end{aligned}$$

where

$$\Xi(\pi_n \overline{\boldsymbol{\xi}}) = \mathsf{X}_{m=1}^{n} [-\overline{\xi}_m, \overline{\xi}_m] \times \mathsf{X}_{m=n+1}^{\infty} \{0\}.$$

Since $\Xi(\pi_n \overline{\boldsymbol{\xi}})$ is compact because $\mathsf{X}_{m=1}^{n} [-\overline{\xi}_m, \overline{\xi}_m]$ is compact in $\mathbb{R}^n$ and therefore $\Xi(\pi_n \overline{\boldsymbol{\xi}})$ is sequentially compact [c.f. Lemma 9.2], it follows now that there exists a finite number of elements $\boldsymbol{\xi}_i$, $i = 1, 2, ..., K$, in $\Xi(\pi_n \overline{\boldsymbol{\xi}}) \subset \Xi(\overline{\boldsymbol{\xi}})$ such that

$$\Xi(\overline{\boldsymbol{\xi}}) \subset \cup_{i=1}^{K} \{\boldsymbol{\xi} \in \Xi : ||\boldsymbol{\xi} - \boldsymbol{\xi}_i|| < \varepsilon\}.$$

Thus, $\Xi(\overline{\boldsymbol{\xi}})$ is totally bounded.

### 9.9.2 Lemma 9.4

First note that $\Xi_\ell(M) \subset \mathsf{X}_{m=1}^{\infty} [-M.m^{-\ell}, M.m^{-\ell}]$, which by Lemma 9.3 is compact with respect to the pseudo-Euclidean metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||$. Then by Lemma 9.1, $\Xi_\ell(M)$ is totally bounded with respect to the pseudo-Euclidean metric $||\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2||$. Next, let $\boldsymbol{\xi}_n = \{\xi_{n,m}\}_{m=1}^{\infty}$ be a Cauchy sequence in $\Xi_\ell(M)$, which of course is also a Cauchy sequence in $\mathsf{X}_{m=1}^{\infty} [-M.m^{-\ell}, M.m^{-\ell}]$. Then there exists a $\underline{\boldsymbol{\xi}} = \{\underline{\xi}_m\}_{m=1}^{\infty} \in \mathsf{X}_{m=1}^{\infty} [-M.m^{-\ell}, M.m^{-\ell}]$ such that for each $m \in \mathbb{N}$, $\lim_{n\to\infty} \xi_{n,m} = \underline{\xi}_m$. Consequently, for each $k \in \mathbb{N}$,

$$M \geq \lim_{n\to\infty} \sum_{m=1}^{k} m^\ell |\xi_{n,m}| = \sum_{m=1}^{k} m^\ell |\underline{\xi}_m|.$$

Letting $k \to \infty$, it follows now that $\underline{\boldsymbol{\xi}} \in \Xi_\ell(M)$. Thus, $\Xi_\ell(M)$ is complete with respect to the pseudo-Euclidean metric, hence $\Xi_\ell(M)$ is compact with respect to this metric.

It follows now from Lemma 9.2 that for an arbitrary sequence $\boldsymbol{\xi}_n = \{\xi_{n,m}\}_{m=1}^{\infty} \in \Xi_\ell(M)$ there exist a subsequence $n_k$ and an element $\underline{\boldsymbol{\xi}} = \{\underline{\xi}_m\}_{m=1}^{\infty} \in \Xi_\ell(M)$ such that $\lim_{k\to\infty} ||\boldsymbol{\xi}_{n_k} - \underline{\boldsymbol{\xi}}|| = 0$, which trivially implies that

$$\lim_{k\to\infty} \sup_{m\in\mathbb{N}} |\xi_{n_k,m} - \underline{\xi}_m| = 0.$$

To show that $\lim_{k\to\infty} ||\boldsymbol{\xi}_{n_k} - \underline{\boldsymbol{\xi}}||_\ell = 0$ as well, note that for any $L \in \mathbb{N}$ we have

$$\begin{aligned}
\limsup_{k\to\infty} \sum_{m=1}^{\infty} m^\ell |\xi_{n_k,m} - \underline{\xi}_m| &= \limsup_{k\to\infty} \sum_{m=1}^{L} m^\ell |\xi_{n_k,m} - \underline{\xi}_m| \\
&\quad + \limsup_{k\to\infty} \sum_{m=1+L}^{\infty} m^\ell |\xi_{n_k,m} - \underline{\xi}_m| \\
&= \limsup_{k\to\infty} \sum_{m=1+L}^{\infty} m^\ell |\xi_{n_k,m} - \underline{\xi}_m|.
\end{aligned}$$

Clearly, the latter "lim sup" is invariant for $L$.

Next, denote $S_k(L) = \sum_{m=1+L}^{\infty} m^\ell |\xi_{n_k,m} - \underline{\xi}_m|$ and let $\limsup_{k\to\infty} S_k(L) = \eta$, which does not depend on $L$. Then by the definition of "lim sup",

$$\begin{aligned}
\eta &= \inf_{L\in\mathbb{N}} \left( \limsup_{k\to\infty} S_k(L) \right) = \inf_{L\in\mathbb{N}} \inf_{s\in\mathbb{N}} \sup_{k\geq s} S_k(L) \\
&\leq \inf_{L\in\mathbb{N}} \sup_{k\geq s} S_k(L) \quad \text{for all } s \in \mathbb{N}.
\end{aligned}$$

Since $\sup_{k\geq s} S_k(L)$ is decreasing in $L$ we have

$$\inf_{L\in\mathbb{N}} \sup_{k\geq s} S_k(L) = \sup_{k\geq s} S_k(\infty) = \sup_{k\geq s} \lim_{L\to\infty} S_k(L),$$

hence, for an arbitrary $s \in \mathbb{N}$,

$$\begin{aligned}
\eta &\leq \sup_{k\geq s} \lim_{L\to\infty} S_k(L) \\
&= \sup_{k\geq s} \left( \lim_{L\to\infty} \sum_{m=1+L}^{\infty} m^\ell |\xi_{n_k,m} - \underline{\xi}_m| \right) = 0.
\end{aligned}$$

The latter follows from the fact that $\sum_{m=1}^{\infty} m^\ell |\xi_{n_k,m} - \underline{\xi}_m| \leq 2M$. Consequently, $\lim_{k\to\infty} ||\boldsymbol{\xi}_{n_k} - \underline{\boldsymbol{\xi}}||_\ell = 0$. This completes the proof of Lemma 9.4.

# Chapter 10

# Asymptotic normality under high-level conditions

## 10.1 Introduction

As is well known, the asymptotic normality result (9.26) is equivalent to

$$\sqrt{N}\tau'(\widehat{\theta}_{n_N} - \theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \tau'\Sigma\tau) \tag{10.1}$$

for an arbitrary $\tau \in \mathbb{R}^d$.

The latter can be written in terms of $\widehat{\boldsymbol{\xi}}_{n_N}$ and $\boldsymbol{\xi}^0$ as follows. First, observe that the metric space $\Xi_\ell(M)$ in Theorem 9.3 is contained in the pseudo-Euclidean space

$$\Xi = \left\{ \boldsymbol{\xi} = \{\xi_m\}_{m=1}^{\infty} : \sum_{m=1}^{\infty} \xi_m^2 < \infty \right\}.$$

Moreover, it has been shown in Theorem 2.7 that if we endow this space with the pseudo-Euclidean innerproduct

$$\langle \boldsymbol{\xi}_1, \boldsymbol{\xi}_2 \rangle_{\mathrm{E}} = \sum_{m=1}^{\infty} \xi_{1,m}\xi_{2,m}, \text{ where } \boldsymbol{\xi}_i = \{\xi_{i,m}\}_{m=1}^{\infty},$$

and associated norm $||\boldsymbol{\xi}||_{\mathrm{E}} = \sqrt{\langle \boldsymbol{\xi}, \boldsymbol{\xi} \rangle_{\mathrm{E}}}$ and metric it becomes a Hilbert space. Furthermore, note that the result of Theorem 9.3 implies that $\mathrm{plim}_{N\to\infty} ||\widehat{\boldsymbol{\xi}}_{n_N} - \boldsymbol{\xi}^0||_{\mathrm{E}} = 0$ as well.

Denoting

$$\rho(\boldsymbol{\xi}) = \langle \pi_d \boldsymbol{\tau}, \boldsymbol{\xi} \rangle_{\mathrm{E}}\,, \text{ where } \boldsymbol{\tau} \in \Xi \text{ is arbitrary,} \tag{10.2}$$

which is clearly linear and continuous in $\boldsymbol{\xi}$, and with $\tau$ the vector of the first $d$ elements of $\boldsymbol{\tau}$, the result (10.1) now reads

$$\sqrt{N}\rho(\widehat{\boldsymbol{\xi}}_{n_N} - \boldsymbol{\xi}^0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \tau'\Sigma\tau). \tag{10.3}$$

In his seminal paper, Shen (1997) constructs a Hilbert space $\mathcal{H}$, say, with innerproduct $\langle ., . \rangle$ and associated norm $||.||$ and metric, such that $\rho(\boldsymbol{\xi})$ is also linear and continuous on $\mathcal{H}$, where the innerproduct $\langle ., . \rangle$ is chosen such that the matrix $\Sigma$ becomes the asymptotically efficient variance matrix.

Shen's approach hinges on the Riesz representation theorem for linear functionals and the notion of directional derivatives, which I will discuss in the next sections.

## 10.2 The Riesz representation theorem

Let $\mathcal{H}$ be a Hilbert space and let $\rho(\boldsymbol{\xi})$ be a continuous real valued linear functional on $\mathcal{H}$, i.e.,

1. $\rho : \mathcal{H} \to \mathbb{R}$;
2. For all $\varepsilon > 0$ and $x \in \mathcal{H}$ there exists a $\delta > 0$, possibly depending on $\boldsymbol{\xi}$, such that $|\rho(\boldsymbol{\xi}) - \rho(\boldsymbol{\varsigma})| < \varepsilon$ if $||\boldsymbol{\xi} - \boldsymbol{\varsigma}|| < \delta$, with $\boldsymbol{\varsigma} \in \mathcal{H}$;
3. For all $\boldsymbol{\xi}, \boldsymbol{\varsigma} \in \mathcal{H}$ and all scalars $\alpha$, $\beta \in \mathbb{R}$, $\rho(\alpha\boldsymbol{\xi} + \beta\boldsymbol{\varsigma}) = \alpha\rho(\boldsymbol{\xi}) + \beta\rho(\boldsymbol{\varsigma})$.

The Riesz representation theorem states that

**Theorem 10.1.** *There exists a unique $\boldsymbol{v} \in \mathcal{H}$ such that*

$$\rho(\boldsymbol{\xi}) = \langle \boldsymbol{\xi}, \boldsymbol{v} \rangle\,, \tag{10.4}$$

*and*

$$||\rho||_{\sup} \stackrel{\text{def.}}{=} \sup_{\boldsymbol{\xi} \in \mathcal{H}:\ ||\boldsymbol{\xi}||=1} |\rho(\boldsymbol{\xi})| = ||\boldsymbol{v}||. \tag{10.5}$$

## 10.3 Directional derivatives and the Hilbert space they imply

Let $\boldsymbol{\upsilon} = \mathsf{X}_{m=1}^{\infty}\{\upsilon_m\} \in \Xi - \boldsymbol{\xi}^0$, where the latter is the space of elements $\boldsymbol{\upsilon} = \boldsymbol{\xi} - \boldsymbol{\xi}^0$ with $\boldsymbol{\xi} \in \Xi$, and let $\varphi(Z, \boldsymbol{\xi})$ be the log-likelihood function. Suppose that the directional derivative

$$\lim_{t \to 0} \frac{\varphi(Z, \boldsymbol{\xi}^0 + t.\boldsymbol{\upsilon}) - \varphi(Z, \boldsymbol{\xi}^0)}{t} \stackrel{\text{def.}}{=} \varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{\upsilon}]$$

is well-defined and is linear in $\boldsymbol{\upsilon}$. In particular, with $\pi_n$ the truncation operator and assuming that $\varphi(Z, \boldsymbol{\xi})$ is a.s. differentiable in all the components of $\boldsymbol{\xi}^0$, it follows trivially that

$$\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{\upsilon}] = \sum_{m=1}^{n} \upsilon_m \nabla_m \varphi(Z, \boldsymbol{\xi}^0), \tag{10.6}$$

where $\nabla_m$ indicates the partial derivative to component $\xi_m$ of $\boldsymbol{\xi}$.

Now let $\mathcal{V}$ be the space of elements $\boldsymbol{\upsilon} = \mathsf{X}_{m=1}^{\infty}\{\upsilon_m\} \in \Xi - \boldsymbol{\xi}^0$ for which

$$E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{\upsilon}]\right)^2\right] < \infty, \tag{10.7}$$

$$\lim_{n \to \infty} E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{\upsilon}] - \varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{\upsilon}]\right)^2\right] = 0. \tag{10.8}$$

Note that without loss of generality we may replace condition (10.7) by

$$\forall n \in \mathbb{N},\ E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{\upsilon}]\right)^2\right] < \infty, \tag{10.9}$$

because then condition (10.8) implies condition (10.7), and vice versa, conditions (10.7) and (10.8) imply condition (10.9). Moreover, note that by the Cauchy-Schwarz and Chebyshev inequalities the conditions (10.7) and (10.8) imply

$$\lim_{n \to \infty} E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{\upsilon}]\right)^2\right] = E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{\upsilon}]\right)^2\right],$$

$$\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{\upsilon}] = \operatorname*{plim}_{n \to \infty} \sum_{m=1}^{n} \upsilon_m \nabla_m \varphi(Z, \boldsymbol{\xi}^0).$$

The latter implies that $\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{\upsilon}]$ is a.s. linear in $\boldsymbol{\upsilon}$.

For $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathcal{V}$, define the innerproduct

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle = E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}_1]\right)\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}_2]\right)\right] \tag{10.10}$$

with associated norm $||\boldsymbol{v}|| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$ and metric $||\boldsymbol{v}_1 - \boldsymbol{v}_2||$. Then

**Theorem 10.2**. *The closure $\overline{\mathcal{V}}$ of the space $\mathcal{V}$ of elements $\boldsymbol{v} \in \Xi - \boldsymbol{\xi}^0$ satisfying the conditions* (10.7) *and* (10.8), *endowed with the innerproduct* (10.10) *and associated norm and metric, is a Hilbert space.*

The space $\overline{\mathcal{V}}$ is now just the Hilbert space $\mathcal{H}$ mentioned in the previous sections.

## 10.4 Shen's asymptotic normality approach

By the first-order condition for a maximum of $E[\varphi(Z, \boldsymbol{\xi})]$ in $\boldsymbol{\xi} = \boldsymbol{\xi}^0$,and under regularity conditions related to the dominated convergence theorem, we have for all $m \in \mathbb{N}$,

$$E[\nabla_m \varphi(Z, \boldsymbol{\xi}^0)] = \nabla_m E[\varphi(Z, \boldsymbol{\xi}^0)] = 0$$

hence

$$E\left(\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{v}]\right) = 0 \tag{10.11}$$

so that by (10.8),

$$\begin{aligned}
\left|E\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}]\right)\right| &= \left|E\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}] - \varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{v}]\right)\right| \\
&\leq E\left[\left|\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}] - \varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{v}]\right|\right] \\
&\leq \sqrt{E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}] - \varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{v}]\right)^2\right]} \\
&\to 0.
\end{aligned}$$

Thus,

$$E\left(\varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}]\right) = 0. \tag{10.12}$$

As before, let $\widehat{Q}_N(\boldsymbol{\xi}) = (1/N)\sum_{j=1}^N \varphi(Z_j, \boldsymbol{\xi})$ be the log-likelihood function divided by $N$, and denote for $\boldsymbol{v} \in \overline{\mathcal{V}}$,

$$\widehat{Q}'_N(\boldsymbol{\xi}^0)[\boldsymbol{v}] = \lim_{t \to 0} \frac{\widehat{Q}_N(\boldsymbol{\xi}^0 + t.\boldsymbol{v}) - \widehat{Q}_N(\boldsymbol{\xi}^0)}{t} = \frac{1}{N}\sum_{j=1}^N \varphi'(Z_j, \boldsymbol{\xi}^0)[\boldsymbol{v}].$$

Then it follows from (10.12) and the standard central limit theorem, given standard regularity conditions, that

$$\sqrt{N}\widehat{Q}_N'(\boldsymbol{\xi}^0)[\boldsymbol{\upsilon}] \xrightarrow{\text{d}} \mathcal{N}(0,||\boldsymbol{\upsilon}||^2), \tag{10.13}$$

where by (10.6) and (10.11),

$$\begin{aligned} ||\boldsymbol{\upsilon}||^2 &= \lim_{n\to\infty}\sum_{m=1}^{n}\sum_{k=1}^{n}\upsilon_m\upsilon_k E\left[\nabla_m\varphi(Z,\boldsymbol{\xi}^0)\nabla_k\varphi(Z,\boldsymbol{\xi}^0)\right] \\ &= \lim_{n\to\infty}(\pi^n\boldsymbol{\upsilon})'\operatorname{Var}\left(\left(\nabla_1\varphi(Z,\boldsymbol{\xi}^0),...,\nabla_n\varphi(Z,\boldsymbol{\xi}^0)\right)'\right)(\pi^n\boldsymbol{\upsilon}), \end{aligned} \tag{10.14}$$

with $\pi^n\boldsymbol{\upsilon}$ the vector of the first $n$ elements of $\boldsymbol{\upsilon}$.

Since $\rho(\boldsymbol{\xi})$ is a.s. continuous and linear on $\Xi$ it is also continuous and linear on $\overline{\mathcal{V}}$. Then by the Riesz representation theorem there exists a unique $\boldsymbol{\upsilon}^* \in \overline{\mathcal{V}}$ such that

$$\rho(\boldsymbol{\xi}) = \langle\boldsymbol{\xi},\boldsymbol{\upsilon}^*\rangle\,,$$

hence

$$\sqrt{N}\rho(\widehat{\boldsymbol{\xi}}_{n_N}-\boldsymbol{\xi}^0) = \sqrt{N}\left\langle\widehat{\boldsymbol{\xi}}_{n_N}-\boldsymbol{\xi}^0,\boldsymbol{\upsilon}^*\right\rangle.$$

Shen (1997) now establishes high-level conditions such that

$$\sqrt{N}\left\langle\widehat{\boldsymbol{\xi}}_{n_N}-\boldsymbol{\xi}^0,\boldsymbol{\upsilon}^*\right\rangle = \sqrt{N}\widehat{Q}_N'(\boldsymbol{\xi}^0)[\boldsymbol{\upsilon}^*]+o_p(1), \tag{10.15}$$

so that by (10.1), (10.2) and (10.13),

$$\sqrt{N}(\tau'\widehat{\theta}_{n_N}-\tau'\theta_0) = \sqrt{N}\left\langle\widehat{\boldsymbol{\xi}}_{n_N}-\boldsymbol{\xi}^0,\boldsymbol{\upsilon}^*\right\rangle \xrightarrow{\text{d}} \mathcal{N}(0,||\boldsymbol{\upsilon}^*||^2). \tag{10.16}$$

As to the asymptotic variance $||\boldsymbol{\upsilon}^*||^2$, recall from Theorem 10.1 and (10.2) that

$$\begin{aligned} ||\boldsymbol{\upsilon}^*|| &= \sup_{||\boldsymbol{\xi}||=1}\rho(\boldsymbol{\xi}) = \sup_{||\boldsymbol{\xi}||=1}\langle\pi_d\boldsymbol{\tau},\boldsymbol{\xi}\rangle_{\text{E}} = \sup_{||\boldsymbol{\xi}||=1}\langle\pi_d\boldsymbol{\tau},\pi_d\boldsymbol{\xi}\rangle_{\text{E}} \\ &= \sup_{||\boldsymbol{\xi}||=1}\tau'(\pi^d\boldsymbol{\xi}) \end{aligned}$$

where $\pi^d\boldsymbol{\xi}$ is the vector of the first $d$ elements of $\boldsymbol{\xi}$. Moreover, it is easy to see that

$$\sup_{||\boldsymbol{\xi}||=1}\tau'(\pi^d\boldsymbol{\xi}) = \lim_{n\to\infty}\sup_{||\pi_n\boldsymbol{\xi}||=1}\tau'(\pi^d\boldsymbol{\xi})$$

Furthermore, for $n > d$ we can write

$$\tau'(\pi^d \boldsymbol{\xi}) = (\tau', 0'_{n-d})(\pi^n \boldsymbol{\xi})$$

Now denote

$$B_n = \mathrm{Var}\left(\left(\nabla_1 \varphi(Z, \boldsymbol{\xi}^0), ..., \nabla_n \varphi(Z, \boldsymbol{\xi}^0)\right)'\right) \tag{10.17}$$

and assume that $B_n$ is finite and nonsingular for each $n \in \mathbb{N}$. Moreover, note that similar to (10.14) it follows that

$$||\pi_n \boldsymbol{\xi}||^2 = (\pi^n \boldsymbol{\xi})' B_n (\pi^n \boldsymbol{\xi}).$$

Thus, $\arg\max_{||\pi_n \boldsymbol{\xi}||=1} \tau'(\pi^d \boldsymbol{\xi})$ is the solution of the maximization problem: Maximize $(\tau', 0'_{n-d})x$ to $x \in \mathbb{R}^n$ subject to $x' B_n x = 1$. It is now a standard Lagrangian exercise to verify that

$$||\boldsymbol{\upsilon}^*||^2 = \lim_{n\to\infty} (\tau', 0'_{n-d}) B_n^{-1} \left( \begin{array}{c} \tau \\ 0_{n-d} \end{array} \right).$$

Thus, the matrix $\Sigma$ in (9.26) is

$$\Sigma = \lim_{n\to\infty} (I_d, O_{d\times(n-d)}) B_n^{-1} \left( \begin{array}{c} I_d \\ O_{(n-d)\times d} \end{array} \right),$$

provided of course that this limit exists.

## 10.5 Alternative conditions for Shen's results

However, rather than discussing Shen's conditions for (10.15), which are too high-level for my taste, I will set forth alternative conditions that are closer related to my approach in Bierens (2014), mimicking the standard asymptotic normality conditions for finite-dimensional ML models. Also, these conditions will provide more intuition than those in Shen (1997).

The first condition is that

$$\sqrt{N} \widehat{Q}'_N(\widehat{\boldsymbol{\xi}}_{n_N})[\pi_{n_N} \boldsymbol{\upsilon}^*] = o_p(1). \tag{10.18}$$

Recall that

$$\begin{aligned} \widehat{Q}'_N(\widehat{\boldsymbol{\xi}}_{n_N})[\pi_{n_N} \boldsymbol{\upsilon}^*] &= \sum_{m=1}^{n_N} \upsilon_m^* \frac{1}{N} \sum_{j=1}^{N} \nabla_m \varphi(Z, \widehat{\boldsymbol{\xi}}_{n_N}) \\ &= \sum_{m=1}^{n_N} \upsilon_m^* \nabla_m \widehat{Q}_N(\widehat{\boldsymbol{\xi}}_{n_N}), \end{aligned}$$

hence if $\widehat{\boldsymbol{\xi}}_{n_N}$ is an interior point of its sieve space $\Xi_{n_N}$ then by the first-order conditions for a maximum of $\widehat{Q}_N(\boldsymbol{\xi})$ on $\Xi_{n_N}$, $\nabla_m \widehat{Q}_N(\widehat{\boldsymbol{\xi}}_{n_N}) = 0$ for $m = 1, 2, ..., n_N$, so that $\widehat{Q}'_N(\widehat{\boldsymbol{\xi}}_{n_N})[\pi_{n_N}\boldsymbol{\upsilon}^*] = 0$, which implies (10.18). However, as argued in Bierens (2014), in general there is *no guarantee that*

$$\lim_{N\to\infty} \Pr\left[\nabla_m \widehat{Q}_N(\widehat{\boldsymbol{\xi}}_{n_N}) = 0 \text{ for } m = 1, 2, ..., n_N\right] = 1,$$

except if $\Xi$ is compact with $\boldsymbol{\xi}^0$ in its interior, as in the case of Theorem 9.3.

Next, let $\varphi(Z, \boldsymbol{\xi})$ be a.s. twice continuously differentiable in the elements of $\boldsymbol{\xi}$. In particular, denote $\nabla_{k,m}\varphi(Z, \boldsymbol{\xi}) = \partial^2\varphi(Z, \boldsymbol{\xi})/(\partial\xi_k\partial\xi_m)$. Then by the mean value theorem there exists a sequence $\lambda_{m,N} \in [0, 1]$ such that

$$\begin{aligned}
\sqrt{N}\nabla_m \widehat{Q}_N(\widehat{\boldsymbol{\xi}}_{n_N}) &= \sqrt{N}\nabla_m \widehat{Q}_N(\pi_{n_N}\boldsymbol{\xi}^0) \\
&\quad + \sum_{k=1}^{n_N} \nabla_{k,m}\widehat{Q}_N\left(\pi_{n_N}\boldsymbol{\xi}^0 + \lambda_{m,N}\left(\widehat{\boldsymbol{\xi}}_{n_N} - \pi_{n_N}\boldsymbol{\xi}^0\right)\right) \\
&\quad \times\sqrt{N}\left(\widehat{\xi}_{k,n_N} - \xi_k^0\right) \\
&= \sqrt{N}\nabla_m \widehat{Q}_N(\pi_{n_N}\boldsymbol{\xi}^0) \\
&\quad + \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{n_N} \nabla_{k,m}\varphi(Z_j, \boldsymbol{\xi})\Big|_{\boldsymbol{\xi}=\pi_{n_N}\boldsymbol{\xi}^0+\lambda_{m,N}\left(\widehat{\boldsymbol{\xi}}_{n_N}-\pi_{n_N}\boldsymbol{\xi}^0\right)} \\
&\quad \times\sqrt{N}\left(\widehat{\xi}_{k,n_N} - \xi_k^0\right),
\end{aligned}$$

where $\widehat{\xi}_{k,n_N}$ is element $k$ of $\widehat{\boldsymbol{\xi}}_{n_N}$ and $\xi_k^0$ is element $k$ of $\boldsymbol{\xi}^0$, hence

$$\begin{aligned}
&\sqrt{N}\widehat{Q}'_N(\widehat{\boldsymbol{\xi}}_{n_N})[\pi_{n_N}\boldsymbol{\upsilon}^*] = \sqrt{N}\widehat{Q}'_N(\pi_{n_N}\boldsymbol{\xi}^0)[\pi_{n_N}\boldsymbol{\upsilon}^*] \\
&+ \sum_{m=1}^{n_N} \upsilon_m^* \sum_{k=1}^{n_N} \frac{1}{N}\sum_{j=1}^{N} \nabla_{k,m}\varphi(Z_j, \boldsymbol{\xi})|_{\boldsymbol{\xi}=\pi_{n_N}\boldsymbol{\xi}^0+\lambda_{m,N}\left(\widehat{\boldsymbol{\xi}}_{n_N}-\pi_{n_N}\boldsymbol{\xi}^0\right)} \\
&\times\sqrt{N}\left(\widehat{\xi}_{k,n_N} - \xi_k^0\right).
\end{aligned}$$

Now *assume* that uniformly in $k, m \in \mathbb{N}$,

$$\begin{aligned}
&\frac{1}{N}\sum_{j=1}^{N} \nabla_{k,m}\varphi(Z_j, \boldsymbol{\xi})|_{\boldsymbol{\xi}=\pi_{n_N}\boldsymbol{\xi}^0+\lambda_{m,N}\left(\widehat{\boldsymbol{\xi}}_{n_N}-\pi_{n_N}\boldsymbol{\xi}^0\right)} \sqrt{N}\left(\widehat{\xi}_{k,N} - \xi_k^0\right) \\
&-E\left[\nabla_{k,m}\varphi(Z, \boldsymbol{\xi}^0)\right]\sqrt{N}\left(\widehat{\xi}_{k,n_N} - \xi_k^0\right) = o_p(1), \qquad (10.19)
\end{aligned}$$

and that by the standard ML conditions,

$$E\left[\nabla_{k,m}\varphi(Z,\boldsymbol{\xi}^0)\right] = -E\left[\nabla_k\varphi(Z,\boldsymbol{\xi}^0)\nabla_m\varphi(Z,\boldsymbol{\xi}^0)\right]. \qquad (10.20)$$

Then

$$\begin{aligned} o_p(1) &= \sqrt{N}\widehat{Q}'_N(\widehat{\boldsymbol{\xi}}_{n_N})[\pi_{n_N}\boldsymbol{\upsilon}^*] \\ &= \sqrt{N}\widehat{Q}'_N(\pi_{n_N}\boldsymbol{\xi}^0)[\pi_{n_N}\boldsymbol{\upsilon}^*] \\ &\quad -\sum_{m=1}^{n_N}\upsilon_m^*\sum_{k=1}^{n_N}E\left[\nabla_k\varphi(Z,\boldsymbol{\xi}^0)\nabla_m\varphi(Z,\boldsymbol{\xi}^0)\right] \\ &\quad \times\sqrt{N}\left(\xi_{k,n_N}-\xi_k^0\right)+o_p(1) \\ &= \sqrt{N}\widehat{Q}'_N(\pi_{n_N}\boldsymbol{\xi}^0)[\pi_{n_N}\boldsymbol{\upsilon}^*] - \sqrt{N}\left\langle\pi_{n_N}\boldsymbol{\upsilon}^*,\widehat{\boldsymbol{\xi}}_{n_N}-\pi_{n_N}\boldsymbol{\xi}^0\right\rangle \\ &\quad +o_p(1), \end{aligned}$$

where the latter follows from the definition of the innerproduct (10.10).

Finally, choose $n_N$ such that

$$\sqrt{N}\widehat{Q}'_N(\pi_{n_N}\boldsymbol{\xi}^0)[\pi_{n_N}\boldsymbol{\upsilon}^*] = \sqrt{N}\widehat{Q}'_N(\boldsymbol{\xi}^0)[\boldsymbol{\upsilon}^*]+o_p(1) \qquad (10.21)$$

and

$$\sqrt{N}\left\langle\pi_{n_N}\boldsymbol{\upsilon}^*,\widehat{\boldsymbol{\xi}}_{n_N}-\pi_{n_N}\boldsymbol{\xi}^0\right\rangle = \sqrt{N}\left\langle\boldsymbol{\upsilon}^*,\widehat{\boldsymbol{\xi}}_{n_N}-\boldsymbol{\xi}^0\right\rangle+o_p(1). \qquad (10.22)$$

Then Shen's result (10.16) follows.

Of course, for specific cases of $\varphi(Z,\boldsymbol{\xi})$ these high-level conditions can be broken down into more primitive conditions, but I will postpone that until I have discussed my approach in Bierens (2014) for the SNP Tobit model.

## 10.6 Proofs

### 10.6.1 Theorem 10.1

Suppose that there exist two distinct elements $v_1, v_2 \in \mathcal{H}$ such (10.4) holds. Then $\langle x, v_1 - v_2\rangle = 0$, hence for $x = v_1 - v_2$, $||v_1 - v_2||^2 = \langle v_1 - v_2, v_1 - v_2\rangle = 0$. Therefore, if (10.4) is true then $v$ is unique.

Without loss of generality we may assume that $\varphi(x)$ is not identically zero, as otherwise (10.4) holds for $v = 0$. Denote

$$\mathcal{M} = \{x \in \mathcal{H} : \varphi(x) = 0\}.$$

Clearly, $\mathcal{M}$ is a linear subspace of $\mathcal{H}$ : $x, y \in \mathcal{M}$ and $\alpha$, $\beta \in \mathbb{R}$ imply $\alpha x+\beta y \in \mathcal{M}$. Moreover, $\mathcal{M}$ is closed. To see this, let $x^*$ be a point of closure of $\mathcal{M}$. Then there exists a sequence $x_n \in \mathcal{M}$ such that $\lim_n ||x_n - x^*|| = 0$, hence by the continuity of $\varphi(x)$, $|\varphi(x^*)| = \lim_{n\to\infty} |\varphi(x^*) - \varphi(x_n)| = 0$, and thus $x^* \in \mathcal{M}$. Consequently, $\mathcal{M}$ is a Hilbert space itself.

Let $\mathcal{M}^\perp$ be the orthogonal complement of $\mathcal{M}$, i.e.,

$$\mathcal{M}^\perp = \{u \in \mathcal{H} : \langle x, u\rangle = 0 \text{ for all } x \in \mathcal{M}\},$$

and note that $\mathcal{M}^\perp \neq \{0\}$ because otherwise $\varphi(x)$ is identically zero. Then by the projection theorem, each $z \in \mathcal{H}$ can be written as $z = x + y$, where $x \in \mathcal{M}$ and $y \in \mathcal{M}^\perp$.

Now pick a nonzero element $y \in \mathcal{M}^\perp$, and let $x \in \mathcal{H}$ be arbitrary but fixed. Without loss of generality we may assume that $\varphi(y) = 1$. Write

$$x = (x - \varphi(x).y) + \varphi(x).y$$

and note that $\varphi(x - \varphi(x).y) = \varphi(x) - \varphi(x).\varphi(y) = \varphi(x) - \varphi(x) = 0$, hence $z = (x - \varphi(x).y) \in \mathcal{M}$, whereas obviously $\varphi(x).y \in \mathcal{M}^\perp$. Then

$$\langle x, y\rangle = \langle z, y\rangle + \varphi(x).\langle y, y\rangle = \varphi(x).||y||^2,$$

hence (10.4) holds for $v = ||y||^{-2}y$.

To prove (10.5), let $||x|| = 1$. Then by (10.4) and the Cauchy-Schwarz inequality $|\varphi(x)| = |\langle x, v\rangle| \leq ||x||.||v|| = ||v||$, hence $||\varphi||_{\sup} \leq ||v||$. On the other hand, $||\varphi||_{\sup} \geq |\varphi(||v||^{-1}v)| = ||v||^{-1}\langle v, v\rangle = ||v||$. Hence, $||\varphi||_{\sup} = ||v||$.

### 10.6.2 Theorem 10.2

Note first that by condition (10.8),

$$\begin{aligned} 0 &= \lim_{n\to\infty} E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{v}] - \varphi'(Z, \boldsymbol{\xi}^0)[\boldsymbol{v}]\right)^2\right] \\ &= \lim_{n\to\infty} \{\langle \pi_n \boldsymbol{v}, \pi_n \boldsymbol{v}\rangle - 2\langle \pi_n \boldsymbol{v}, \boldsymbol{v}\rangle + \langle \boldsymbol{v}, \boldsymbol{v}\rangle\} \\ &= \lim_{n\to\infty} ||\boldsymbol{v} - \pi_n \boldsymbol{v}||^2 \end{aligned}$$

for $\boldsymbol{v} \in \mathcal{V}$, hence for an arbitrary $\varepsilon > 0$ there exists an $n$ such that

$$\begin{aligned} \varepsilon &> ||\boldsymbol{v} - \pi_n \boldsymbol{v}||^2 = ||\pi_n \boldsymbol{v}||^2 - 2\langle \pi_n \boldsymbol{v}, \boldsymbol{v}\rangle + ||\boldsymbol{v}||^2 \\ &= ||\boldsymbol{v}||^2 - ||\pi_n \boldsymbol{v}||^2, \end{aligned}$$

so that $||\boldsymbol{v}|| < \sqrt{\varepsilon + ||\pi_n \boldsymbol{v}||^2} < \infty$. Secondly, note that by the definition of closure, for $\boldsymbol{v}_* \in \overline{\mathcal{V}} \backslash \mathcal{V}$ and arbitrary $\varepsilon > 0$ there exists a $\boldsymbol{v} \in \mathcal{V}$ such that $||\boldsymbol{v}_* - \boldsymbol{v}|| < \varepsilon$, hence $||\boldsymbol{v}_*|| \leq ||\boldsymbol{v}_* - \boldsymbol{v}|| + ||\boldsymbol{v}|| < \infty$. Thus, for all $\boldsymbol{v} \in \overline{\mathcal{V}}$, $||\boldsymbol{v}|| < \infty$.

Next, let $\boldsymbol{v}_m$ be a Cauchy sequence in $\overline{\mathcal{V}}$, i.e., $\lim_{\min(k,m)\to\infty} ||\boldsymbol{v}_m - \boldsymbol{v}_k|| = 0$. Then for each fixed $n \in \mathbb{N}$, $\pi_n \boldsymbol{v}_m$ is a Cauchy sequence in $\overline{\mathcal{V}}$, and therefore the vectors $(v_{1,m}, ..., v_{n,m})'$ of the first $n$ elements of $\pi_n \boldsymbol{v}_m$ are Cauchy sequences in $\mathbb{R}^n$. Then by Theorem 2.2,

$$\lim_{m\to\infty} (v_{1,m}, ..., v_{n,m})' = (v_1, ..., v_n)' \in \mathbb{R}^n,$$

hence, for each $n \in \mathbb{N}$ there exists a $\boldsymbol{v} = \{v_k\}_{k=1}^{\infty}$ such that

$$\lim_{m\to\infty} ||\pi_n \boldsymbol{v}_m - \pi_n \boldsymbol{v}|| = 0,$$

provided that $\pi_n \boldsymbol{v} \in \overline{\mathcal{V}}$. Since for all $n, m \in \mathbb{N}$, $\pi_n \boldsymbol{v}_m \in \overline{\mathcal{V}}$, we must have, by the same argument as in the proof of Theorem 2.2, that for all $n \in \mathbb{N}$, $\pi_n \boldsymbol{v} \in \overline{\mathcal{V}}$, hence by condition (10.7), $E\left[\left(\varphi'(Z, \boldsymbol{\xi}^0)[\pi_n \boldsymbol{v}]\right)^2\right] < \infty$ for all $n \in \mathbb{N}$, which is just condition (10.9). Consequently, it follows from condition (10.8) that $\boldsymbol{v} \in \overline{\mathcal{V}}$.

Now for an arbitrary $\varepsilon > 0$ there exists an $m_n \in \mathbb{N}$ such that $||\pi_n \boldsymbol{v}_m - \pi_n \boldsymbol{v}|| < \varepsilon$ if $m \geq m_n$, and by the Cauchy property there exists an $\ell \in \mathbb{N}$ such that $||\boldsymbol{v}_m - \boldsymbol{v}_k|| < \varepsilon$ if $\min(m, k) \geq \ell$. Hence for fixed $k \geq \ell$ and $m \geq \max(\ell, m_n)$,

$$\begin{aligned}
||\boldsymbol{v}_m - \boldsymbol{v}|| &\leq ||\pi_n \boldsymbol{v}_m - \pi_n \boldsymbol{v}|| + ||\boldsymbol{v}_m - \pi_n \boldsymbol{v}_m|| \\
&\quad + ||\boldsymbol{v} - \pi_n \boldsymbol{v}|| \\
&\leq ||\pi_n \boldsymbol{v}_m - \pi_n \boldsymbol{v}|| + ||\boldsymbol{v}_m - \boldsymbol{v}_k|| \\
&\quad + ||\boldsymbol{v}_k - \pi_n \boldsymbol{v}_m|| + ||\boldsymbol{v} - \pi_n \boldsymbol{v}|| \\
&\leq ||\pi_n \boldsymbol{v}_m - \pi_n \boldsymbol{v}|| + ||\boldsymbol{v}_m - \boldsymbol{v}_k|| \\
&\quad + ||\pi_n \boldsymbol{v}_k - \pi_n \boldsymbol{v}_m|| + ||\boldsymbol{v}_k - \pi_n \boldsymbol{v}_k|| \\
&\quad + ||\boldsymbol{v} - \pi_n \boldsymbol{v}|| \\
&\leq ||\pi_n \boldsymbol{v}_m - \pi_n \boldsymbol{v}|| + 2||\boldsymbol{v}_m - \boldsymbol{v}_k|| \\
&\quad + ||\boldsymbol{v}_k - \pi_n \boldsymbol{v}_k|| + ||\boldsymbol{v} - \pi_n \boldsymbol{v}|| \\
&\leq 3\varepsilon + ||\boldsymbol{v}_k - \pi_n \boldsymbol{v}_k|| + ||\boldsymbol{v} - \pi_n \boldsymbol{v}||
\end{aligned}$$

and thus

$$\limsup_{m\to\infty} ||\boldsymbol{v}_m - \boldsymbol{v}|| \leq 3\varepsilon + ||\boldsymbol{v}_k - \pi_n \boldsymbol{v}_k|| + ||\boldsymbol{v} - \pi_n \boldsymbol{v}||.$$

Letting $n \to \infty$ it follows now that $\limsup_{m\to\infty} ||\boldsymbol{v}_m - \boldsymbol{v}|| \leq 3\varepsilon$, hence by the arbitrariness of $\varepsilon > 0$ we have $\lim_{m\to\infty} ||\boldsymbol{v}_m - \boldsymbol{v}|| = 0$.

# Chapter 11

# To be continued

## References[1]


Adams, R. A. & J. J..F. Fournier (2003), *Sobolev Spaces*, Academic Press.

Anderson, T. W. (1994), *The Statistical Analysis of Time Series*, Wiley.

Bahadur, R.R. (1967), "Rates of Convergence of Estimates and Test Statistics", *Annals of Mathematical Statistics* 38, 303-324.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov & J.A. Wellner (1998), *Efficient and Adaptive Estimation for Semiparametric Models*, Springer.

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

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. (2014), "Consistency and Asymptotic Normality of Sieve Estimators Under Low-Level Conditions", *Econometric Theory* 30, 1021-1076.


[1] Some of the references below are not yet referred to in the text above. They will be in due course.

Bierens, H.J. (2017), *Econometric Model Specification: Consistent Model Specification Tests and Semi-Nonparametric Modeling and Inference*, World Scientific Publishers.

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. & H. Pott-Buter (1990), "Specification of Engel Curves by Nonparametric Regression (with discussion)", *Econometric Reviews* 9, 123-184.

Bierens, H. J. & H. Song (2012): "Semi-Nonparametric Estimation of Independently and Identically Repeated First-Price Auctions via an Integrated Simulated Moments Method", *Journal of Econometrics* 168, 108-119.

Chen, X. (2007), "Large Sample Sieve Estimation of Semi-Nonparametric Models". In J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics, Vol. 6,* Ch. 76. Elsevier.

Ding, Y. & B. Nan (2011), "A Sieve M-Theorem for Bundled Parameters in Semiparametric Models, with Application to the Efficient Estimation in a Linear Model for Censored Data", *Annals of Statistic* 39, 3032-3061.

Duncan, G. M. (1986), "A Semi-Parametric Censored Regression Estimator", *Journal of Econometrics* 32, 5-34.

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.

Gallant, A.R. & D.W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica* 55, 363-390.

Hamming, R.W. (1973), *Numerical Methods for Scientists and Engineers.* Dover Publications.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica* 47, 153-161.

Heckman, J. J. & B. Singer (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration

Data", *Econometrica* 52, 271-320.

Jennrich, R.I. (1969), "Asymptotic Properties of Nonlinear Least Squares Estimators", *Annals of Mathematical Statistics* 40, 633-643.

Khan, S.& J. L. Powell (2001), "Two-step Estimation of Semiparametric Censored Regression Models", *Journal of Econometrics* 103, 73-110.

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.

Lewbel, A. & O. Linton (2002), "Nonparametric Censored and Truncated Regression", *Econometrica* 70, 765-779.

Manski, C.F. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics* 27, 313-333.

Manski, C.F. (1988), "Identification of Binary Response Models", *Journal of the American Statistical Association* 83, 729-738.

Olsen, R. (1978), "A Note on the Uniqueness of the Maximum Likelihood Estimator in the Tobit Model", *Econometrica* 46, 1211-1215.

Powell, J. L. (1984), "Least Absolute Deviations Estimation for the Censored Regression Model", *Journal of Econometrics* 25, 303-325.

Powell, J. L. (1986), "Symmetrically Trimmed Least Squares Estimation of Tobit Models", *Econometrica* 54, 1435-1460.

Powell, J. L. (1994), "Estimation of Semiparametric Models". In R. F. Engle and D.L. McFadden (eds), *Handbook of Econometrics, Vol. 4*, Ch. 41. North-Holland.

Powell, J. L. (2008), "Semiparametric Estimation". In S. N. Durlauf and L. E. Blume (eds), *The New Palgrave Dictionary of Economics. Second Edition.* Palgrave Macmillan.

Royden, H. L. (1968), *Real Analysis.* Macmillan.

Shen, X. (1997), "On the Method of Sieves and Penalization", *Annals of Statistics* 25, 2555-2591.

Stewart, M.K. B. (2004), "Semi-Nonparametric Estimation of Extended Ordered Probit Models." *The Stata Journal* 4, 27-39.

Tobin, J. (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica* 26, 24-36.

Young, N. (1988), *An Introduction to Hilbert Space.* Cambridge University Press.

Sims, C.A. (1980), "Macroeconomics and Reality", *Econometrica* 48, 1-48.

van der Vaart, A.W. (1998), *Asymptotic Statistics.* Cambridge University Press.

van der Vaart, A.W. & J.A Wellner (1996). *Weak Convergence and Empirical Processes.* Springer.

Wald, A. (1949), "Note on the Consistency of the Maximum Likelihood Estimate", *Annals of Mathematical Statistics* 20, 595-601.

Wold, H. (1938), *A Study in the Analysis of Stationary Time Series.* Almqvist and Wiksell, Sweden.