Addendum to: Asymptotic Theory of Integrated Conditional Moment Tests

1. Introduction

In this addendum to Bierens and Ploberger (1997) [BP hereafter], I will provide the proof of a slightly updated version of Theorem 1 in BP, and the difference with the related result of Stinchcombe and White (1998) will be explained. Moreover, I will provide the proof of Mercer's theorem (Lemma 1 in BP), including the Hilbert-Schmidt theorem regarding the existence of eigenvalues and eigenfunctions. Finally, it will be shown how to compute bootstrap critical values of the ICM test.

The paper under review is to a large extend the work of Werner Ploberger. In particular, the idea of using Mercer's theorem (Lemma 1) to prove that the ICM test has nontrivial power against $\sqrt{n}$ local alternatives (Theorem 3 and Corollary 1), the results in Sections 4 and 5, and the tightness result in Lemma A.1, are all due to Werner.

It was claimed that Theorem 1 in BP is a straightforward extension of Theorem 1 of Bierens (1982, 1990). However, it appears that the proof of the former Theorem 1 is not as straightforward an extension of the results in Bierens (1982, 1990). Therefore, in this addendum I will give a detailed proof of Theorem 1. Moreover, the second part of Theorem 1 is now compared with the related result in Stinchcombe and White (1998), and the difference between these results will be explained.

As to Lemma 1 in BP, it was claimed that this lemma is a straightforward further elaboration of Mercer's theorem, with a reference to Dunford and Schwartz (1963, p. 1088)), mimicking the properties of eigenvalues and eigenvectors of positive definite symmetric matrices. However, the reference involved is a reference to an exercise. The proper reference should have been Mercer (1909), who derived this result for the case of real valued symmetric positive semi-definite functions on $[a, b] \times [a, b]$, with Lebesgue measure μ. Mercer's (1909) proof builds upon results of David Hilbert and his Ph.D. student Erhard Schmidt in a series of papers published in the period 1904-1908, regarding the existence of eigenvalues and corresponding eigenfunctions of positive semidefinite kernels. Their results are nowadays referred to as the Hilbert-Schmidt theorem. See for example Bernkopf (1966) and Siegmund-Schultze (1986) and the references therein. However, the results in Lemma 1 are still not straightforward extensions of the original Hilbert-Schmidt and Mercer theorems. Therefore, in this addendum to BP I will provide a complete proof of Lemma 1, including proof of the Hilbert-Schmidt theorem.

The proofs in this addendum to BP are based on Hilbert space theory at the level of Bierens (2014), linear algebra at the level of Bierens (2004, Appendix I), and measure and probability theory at the level of Bierens (2004).

2. Theorem 1

Recall that in Theorem 1 in BP,

Assumption 2.1. $X = (X_1, ..., X_k)' \in \mathbb{R}^k$ *is a bounded random vector,*

and

Assumption 2.2. U *is a random variable satisfying* $E[|U|] < \infty$ *and* $\Pr(E[U|X] = 0) < 1$.

The first condition in Assumption 2.2 guarantees that the conditional expectation $E[U|X]$ is well-defined, and takes the form $E[U|X] = g(X)$ a.s., where g is a Borel measurable real function on $\mathbb{R}^k$.[1]

Moreover, it was assumed that

Assumption 2.3. *The weight function* $w(u)$ *is a complex or real valued function that is infinitely many times continuously differentiable in* $u = 0$ *and satisfies the condition that, with* $w_n(u) = (\mathrm{d}/\mathrm{d}u)^n w(u)$,

$$\textit{the set } \{n \in \mathbb{N} : w_n(0) = 0\} \textit{ is finite or empty.} \tag{2.1}$$

Now recall that Theorem 1 has two parts:

1. Given Assumptions 2.1-2.3, for any $\varepsilon > 0$ there exists a $\xi \in \mathbb{R}^k$ such that $E[U.w(\xi' X)] \neq 0$ and $||\xi|| < \varepsilon$.

[1] See for example Bierens (2004, Theorem 3.10, p. 77).

2. If in addition $w(u)$ is a power series on an open interval R_0 of $\mathbb{R}$ with closure containing 0,[2] i.e., $w(u) = \sum_{n=0}^{\infty}(\gamma_n/n!)u^n$, with $\gamma_n = w_n(0)$, $\gamma_0 = w(0)$, then the set $S = \left\{\xi \in \mathbb{R}^k : E[U.w(\xi'X)] = 0 \text{ and } \Pr[\xi'X \in R_0] = 1\right\}$ has Lebesgue measure zero and is nowhere dense.

The reason for assuming that the closure of R_0 contains 0 is that then by part 1 and the boundedness of X there exists a $\xi_0 \in \mathbb{R}^k$ such that $\Pr[\xi_0'X \in R_0] = 1$ and $E[U.w(\xi_0'X)] \neq 0$.

2.1. Theorem 1 revised

As said before, the proof of Theorem 1 in BP is not as straightforward an extension of the results in Bierens (1982, 1990) as claimed. In Bierens (1982) the complex exp(.) function $w(u) = \exp(\mathbf{i}.u) = \sum_{n=0}^{\infty}(\mathbf{i}^n/n!)u^n$ was used, and in Bierens (1990) I used the real exp(.) function $w(u) = \exp(u) = \sum_{n=0}^{\infty}(1/n!)u^n$, and in these cases the proof of Theorem 1 is not too difficult.

However, in the current more general case we need additional conditions which were missing in BP. The first missing condition is that:

Assumption 2.4. *For some $\varepsilon > 0$, $w(u)$ is infinitely many times differentiable on $(-\varepsilon, \varepsilon)$, with higher-order derivatives $w_n(u) = (\mathrm{d}/\mathrm{d}u)^n w(u)$ for $n \in \mathbb{N}$, $w_0(u) = w(u)$, satisfying $\sup_{|u|<\varepsilon}|w_n(u)| < \infty$ for each $n \in \mathbb{N}_0$.*[3]

This condition, together with the condition $E[|U|] < \infty$ and the boundedness of X, is needed to guarantee that, by the dominated convergence theorem,

$$\begin{aligned}(\partial/\partial\xi_1)^{m_1}...(\partial/\partial\xi_k)^{m_k}&E[U.w(\xi'X)] \\ &= E[U(\partial/\partial\xi_1)^{m_1}...(\partial/\partial\xi_k)^{m_k}w(\xi'X)]\end{aligned}$$

for ξ in a small open neighborhood of the origin of $\mathbb{R}^k$, where $m_1, ...m_k$ are arbitrary non-negative integers and ξ_i is component i of ξ.[4]

Another missing condition is that

Assumption 2.5. *If $w(u)$ is chosen such that $w_n(0) = 0$ for a finite number of n's then at least one components of X, say X_1, has to satisfy* $\Pr[X_1 \neq 0] = 1$.

[2] Thus, 0 is either a border point or an interior point of R_0.

[3] Here and in the sequel $\mathbb{N}_0$ denotes the set of non-negative integers: $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$.

[4] Note that $(\partial/\partial\xi_i)^0$ should be interpreted as "don't take the partial derivative to ξ_i."

Thus, if the set in (2.1) is empty then we don't need this condition.

Therefore, part 1 of Theorem 1 in BP should be restated as follows.

Theorem 2.1. *Denote* $N_0(\varepsilon) = \{\xi \in \mathbb{R}^k : ||\xi|| < \varepsilon\}$. *Given Assumptions 2.1-2.5, for every* $\varepsilon > 0$ *there exists a* $\xi_0 \in N_0(\varepsilon)$ *such that* $E[U.w(\xi_0' X)] \neq 0$.

In view of the results in Stinchcombe and White (1998) the condition in part 2 of Theorem 1 in BP that $w(u)$ is a power series on R_0 it too strong. Moreover, Assumption 2.4 implicitly assumes that $0 \in R_0$ rather than that the closure of R_0 contains 0. Stinchcombe and White (1998) require that $w(u)$ is a non-polynomial analytical function. Analytical functions are local power series, and power series are analytical in their radius of convergence. Therefore let us adopt the weaker condition that $w(u)$ is analytical on R_0. Note that by Assumption 2.3, $w(u)$ is non-polynomial.

For notational convenience let us assume that $R_0 = \mathbb{R}$. The results in the case $R_0 = \mathbb{R}$ below can easily be adapted to the case that R_0 is a bounded open interval containing 0. Thus,

Assumption 2.6. *Let* $w(u)$ *be a non-polynomial analytical function on* $\mathbb{R}$.

Recall from the definition of an analytical function that for every $u_0 \in \mathbb{R}$ the function $w(u)$ has an infinite Taylor expansion for any u in a small open neighborhood of u_0, i.e.,

$$w(u) = \sum_{n=0}^{\infty} \frac{w_n(u_0)}{n!}(u - u_0)^n, \tag{2.2}$$

and is non-polynomial if

$$\text{the set } \{n \in \mathbb{N} : w_n(0) \neq 0\} \text{ is infinite.} \tag{2.3}$$

It is not too hard to verify that (2.3) implies that for each $u \in \mathbb{R}$ the set $\{n \in \mathbb{N} : w_n(u) \neq 0\}$ is infinite.

However, in the current setup the stronger condition (2.1) in Assumption 2.3 is necessary. For example, let X be uniformly $[-0.5, 0.5]$ distributed, and let $U = X$. Then with $w(u) = \cos(u)$ it follows by symmetry that

$$E[U.w(\xi X)] = \int_{-1/2}^{1/2} x \cos(\xi x) \mathrm{d}x \equiv 0.$$

The function $w(u) = \cos(u)$ is analytic and non-polynomial on $\mathbb{R}$, but does not satisfy condition (2.1).

The second part of Theorem 1 will now be reformulated as follows.

Theorem 2.2. *If in addition to the conditions of Theorem 2.1 also Assumption 2.6 holds then the set* $S = \{\xi \in \mathbb{R}^k : E[U.w(\xi' X)] = 0\}$ *has Lebesgue measure zero and is nowhere dense in* $\mathbb{R}^k$.

Remark 2.1. The boundedness of X is essential for the result of Theorem 2.2. For example, let X be a random variable with density

$$f(x) = \frac{1 - \cos(a.x)}{\pi a.x^2}, \ a > 0, \ x \in \mathbb{R},$$

and let $U \equiv 1$. Moreover, let $w(u) = \cos(u) + \sin(u)$. Then

$$\begin{aligned} E[U.w(\xi.X)] &= E[w(\xi.X)] = \int_{-\infty}^{\infty} \cos(\xi.x) f(x) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} \exp(\mathbf{i}.\xi x) f(x) \mathrm{d}x = \max\{1 - |\xi|/a, \ 0\}. \end{aligned}$$

C.f. Chung (1974, p.148). Hence, in this case

$$S = \{\xi \in \mathbb{R} : E[U.w(\xi.X)] = 0\} = \mathbb{R} \backslash [-a, a].$$

2.2. Proof of Theorem 2.1

Suppose that for the ε in Assumption 2.4, $E[U.w(\xi' X)] \equiv 0$ on $N_0(\varepsilon)$. Then for all $n \in \mathbb{N}$ and by repeated application of the dominated convergence theorem (which is applicable by Assumption 2.4),

$$\begin{aligned} 0 &\equiv (\partial/\partial\lambda)^n E[U.w(\lambda.\xi' X)]|_{\lambda=0} \\ &= E[U(\partial/\partial\lambda)^n w(\lambda.\xi' X)]|_{\lambda=0} \\ &= E[U(\xi' X)^n w_n(\lambda.\xi' X)]|_{\lambda=0} \\ &= w_n(0).E[U(\xi' X)^n]. \end{aligned} \quad (2.4)$$

Consider first the case that $w_n(0) \neq 0$ for all $n \geq 0$, so that $E[U(\xi' X)^n] = 0$ for all $n \geq 0$. Then

$$E[U \exp(\mathbf{i}.\xi' X)] = \sum_{n=0}^{\infty} \frac{\mathbf{i}^n}{n!} E[U(\xi' X)^n] = 0$$

on $N_0(\varepsilon)$, which by Theorem 1 in Bierens (1982) implies that $\Pr[E[U|X] = 0] = 1$. The latter violates Assumption 2.2.

To prove the case that $w_n(0) = 0$ for a finite number of n's, recall from (2.4) that $E[U(\xi' X)^n] \equiv 0$ on $N_0(\varepsilon)$ whenever $w_n(0) \neq 0$. Taking partial derivatives, it follows that for all nonnegative integers $m_1, ...m_k$ satisfying $\sum_{i=1}^k m_i = n$, with $w_n(0) \neq 0$,

$$(\partial/\partial\xi_1)^{m_1}...(\partial/\partial\xi_k)^{m_k} E[U(\xi' X)^n] = 0, \tag{2.5}$$

The equality (2.5) is equivalent to

$$E\left[U \prod_{i=1}^k X_i^{m_i}\right] = 0 \tag{2.6}$$

for all nonnegative integers $m_1, ...m_k$ satisfying $\sum_{i=1}^k m_i = n$ with $w_n(0) \neq 0$. To see this, consider the case $k = 2$. Then

$$E[U(\xi' X)^n] = \sum_{j=0}^n \binom{n}{j} \xi_1^j \xi_2^{n-j} E[U.X_1^j X_2^{n-j}]$$

hence for $m_1 = m$, $m_2 = n - m$, with $0 \leq m \leq n$,

$$\begin{aligned}
&(\partial/\partial\xi_1)^{m_1}(\partial/\partial\xi_2)^{m_2} E[U(\xi' X)^n] \\
&= \sum_{j=0}^n \binom{n}{j} \left((\partial/\partial\xi_1)^{m_1}\xi_1^j\right) . \left((\partial/\partial\xi_2)^{m_2}\xi_2^{n-j}\right) E[UX_1^j X_2^{n-j}] \\
&= \sum_{j=m_1}^{n-m_2} \binom{n}{j} \left((\partial/\partial\xi_1)^{m_1}\xi_1^j\right) . \left((\partial/\partial\xi_2)^{m_2}\xi_2^{n-j}\right) E[UX_1^j X_2^{n-j}] \\
&= \binom{n}{m} \left((\partial/\partial\xi_1)^{m}\xi_1^m\right) . \left((\partial/\partial\xi_2)^{n-m}\xi_2^{n-m}\right) E[UX_1^m X_2^{n-m}] \\
&= \binom{n}{m} (m!).(n-m)! E[UX_1^m X_2^{n-m}] \\
&= n! E[U.X_1^m X_2^{n-m}].
\end{aligned} \tag{2.7}$$

Since $w_n(0) = 0$ for only a finite number of n's, there exists a finite n_0 such that $w_n(0) \neq 0$ for all $n \geq n_0$. Then (2.6) implies that for all nonnegative integers $m_1, ...m_k$,

$$E\left[U.X_1^{n_0} \prod_{i=1}^k X_i^{m_i}\right] = 0,$$

which by Theorem 2 in Bierens (1982) and Assumption 2.5 implies that

$$\begin{aligned}
1 &= \Pr\left[E[UX_1^{n_0}|X]=0\right] = \Pr\left[X_1^{n_0}E[U|X]=0\right] \\
&= \Pr\left[X_1=0 \text{ and } E[U|X]\neq 0\right] + \Pr\left[X_1\neq 0 \text{ and } E[U|X]=0\right] \\
&\quad + \Pr\left[X_1=0 \text{ and } E[U|X]=0\right] \\
&= \Pr\left[X_1=0 \text{ and } E[U|X]\neq 0\right] + \Pr\left[E[U|X]=0\right] \\
&\leq \Pr[X_1=0] + \Pr\left[E[U|X]=0\right] \\
&= \Pr\left[E[U|X]=0\right].
\end{aligned}$$

Again, this result contradicts Assumption 2.2.

Remark 2.2. In the case $m=n$ the equality (2.7) reads

$$(\partial/\partial\xi_1)^n E[U(\xi' X)^n] = n!E[UX_1^n X_2^0].$$

But what is the interpretation of X_2^0 if $\Pr[X_2=0]\in(0,1)$? At first sight it seems that we cannot interpret X_2^0 as 1 because 0^0 is not defined, so that X_2^0 is not defined with positive probability. However, with $\{\Omega,\mathcal{F},P\}$ the probability space and $A=\{\omega\in\Omega : X_2(\omega)\neq 0\}$, and assuming that $\Pr[X_1\neq 0]=1$, we have

$$\begin{aligned}
E[U(\xi' X)^n] &= E[U(\xi_1 X_1+\xi_2 X_2)^n] \\
&\overset{\text{def.}}{=} \int U(\omega)(\xi_1 X_1(\omega)+\xi_2 X_2(\omega))^n \mathrm{d}P(\omega) \\
&= \int_{\Omega\backslash A} U(\omega)(\xi_1 X_1(\omega)+\xi_2 X_2(\omega))^n \mathrm{d}P(\omega) \\
&+ \int_A U(\omega)(\xi_1 X_1(\omega)+\xi_2 X_2(\omega))^n \mathrm{d}P(\omega) \\
&= \xi_1^n \int_{\Omega\backslash A} U(\omega)X_1(\omega)^n \mathrm{d}P(\omega) \\
&+ \sum_{j=0}^n \binom{n}{j}\xi_1^j\xi_2^{n-j}\int_A U(\omega)X_1(\omega)^j X_2(\omega)^{n-j}\mathrm{d}P(\omega) \\
&= \xi_1^n \int_{\Omega\backslash A} U(\omega)X_1(\omega)^n \mathrm{d}P(\omega) \\
&+ \xi_1^n \int_A U(\omega)X_1(\omega)^0 X_2(\omega)^n \mathrm{d}P(\omega)
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{j=1}^{n}\binom{n}{j}\xi_1^j\xi_2^{n-j}\int_A U(\omega)X_1(\omega)^j X_2(\omega)^{n-j}\mathrm{d}P(\omega)\\
&=\xi_1^n\int_{\Omega\backslash A} U(\omega)X_1(\omega)^n\mathrm{d}P(\omega)+\xi_1^n\int_A U(\omega)X_2(\omega)^n\mathrm{d}P(\omega)\\
&+\sum_{j=1}^{n}\binom{n}{j}\xi_1^j\xi_2^{n-j}\int_A U(\omega)X_1(\omega)^j X_2(\omega)^{n-j}\mathrm{d}P(\omega)\\
&+\sum_{j=1}^{n}\binom{n}{j}\xi_1^j\xi_2^{n-j}\int_{\Omega\backslash A} U(\omega)X_1(\omega)^j X_2(\omega)^{n-j}\mathrm{d}P(\omega),
\end{aligned}
$$

where the second equality follows from the fact that $X_2(\omega)\neq 0$ on A, so that $X_2(\omega)^0 = 1$ on A, whereas $X_2(\omega) = 0$ on $\Omega\backslash A$. Hence,

$$
\int_{\Omega\backslash A} U(\omega)X_1(\omega)^j X_2(\omega)^{n-j}\mathrm{d}P(\omega) = 0
$$

for $j \geq 1$. Thus,

$$
\begin{aligned}
E[U(\xi' X)^n] &= \xi_1^n\int U(\omega)X_1(\omega)^n\mathrm{d}P(\omega)\\
&\quad+\sum_{j=1}^{n}\binom{n}{j}\xi_1^j\xi_2^{n-j}\int U(\omega)X_1(\omega)^j X_2(\omega)^{n-j}\mathrm{d}P(\omega)\\
&= \xi_1^n E[U.X_1^n]+\sum_{j=1}^{n}\binom{n}{j}\xi_1^j\xi_2^{n-j}E\left[U.X_1^j X_2^{n-j}\right].
\end{aligned}
$$

Consequently, inside an expectation we may interpret X_2^0 as 1 even if $\Pr[X_2 = 0] > 0$.

2.3. Proof of Theorem 2.2

Under the conditions of Theorem 2.1 there exists a $\xi_0 \in \mathbb{R}^k$ such that $E[U.w(\xi_0' X)] \neq 0$. For this point, consider the hyper "balloon" centered around ξ_0 with radius 1:

$$
\mathbb{B}^k = \{\xi \in \mathbb{R}^k : ||\xi - \xi_0|| = 1\}. \tag{2.8}
$$

Note that in the case $k = 1$, $\mathbb{B}^1 = \{\xi_0 - 1, \xi_0 + 1\}$ and in the case $k = 2$, $\mathbb{B}^2$ is the circle with center ξ_0 and radius 1.

Any $\xi \in \mathbb{R}^k \backslash \{\xi_0\}$ corresponds uniquely to a $\xi_* \in \mathbb{B}^k$, namely the point of intersection of the line piece connecting ξ_0 and ξ with the set $\mathbb{B}^k$, i.e., $\xi_* = \xi_0 + ||\xi - \xi_0||^{-1}(\xi - \xi_0)$. Conversely, for any $\xi_* \in \mathbb{B}^k$ the line piece originating from ξ_0 in the direction ξ_* towards infinity takes the form $\mathbb{L}(\xi_*) = \bigcup_{\lambda \geq 0} \mathbb{L}(\lambda|\xi_*)$, where for $\lambda \geq 0$,

$$\mathbb{L}(\lambda|\xi_*) = \left\{\xi \in \mathbb{R}^k : \xi = \xi_0 + \lambda(\xi_* - \xi_0)\right\}.$$

Next, denote

$$\begin{aligned} \rho(\lambda|\xi_*) &= E[U.w(\xi' X)] \text{ for } \xi \in \mathbb{L}(\lambda|\xi_*), \\ \Lambda(\xi_*) &= \{\lambda > 0 : \rho(\lambda|\xi_*) = 0\}. \end{aligned}$$

I will show now that for arbitrary $\xi_* \in \mathbb{B}^k$ the set $\Lambda(\xi_*)$ is either empty or countable and nowhere dense in $(0, \infty)$, as follows. Let

$$\lambda_1 = \inf_{\lambda \in \Lambda(\xi_*)} \lambda.$$

which is the smallest λ for which $\rho(\lambda|\xi_*) = 0$. If such a λ_1 does not exists then $\Lambda(\xi_*)$ is empty, and thus has Lebesgue measure zero and is nowhere dense in $(0, \infty)$. Otherwise, suppose that for some small $\varepsilon > 0$, $\rho(\lambda|\xi_*) = 0$ for all $\lambda \in [\lambda_1, \lambda_1 + \varepsilon)$, and let $\xi_1 = \xi_* + \lambda_1(\xi_* - \xi_0)$. Then for all $y \in [0, \varepsilon)$,

$$\begin{aligned} 0 &= E[U.w((\xi_1 + y(\xi_* - \xi_0))' X)] \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} y^n E\left[U.w_n(\xi_1' X).\left(((\xi_* - \xi_0)' X\right)^n\right], \end{aligned}$$

hence for $m = 0, 1, 2, ...,$[5]

$$\begin{aligned} 0 \quad = \quad & (\partial/\partial y)^m E[U.w((\xi_1 + y(\xi_* - \xi_0))' X)]|_{y \downarrow 0} \\ & = (\partial/\partial y)^m \sum_{n=0}^{\infty} \frac{1}{n!} y^n E\left[U.w_n(\xi_1' X).\left(((\xi_* - \xi_0)' X\right)^n\right] \Bigg|_{y \downarrow 0} \\ & = \sum_{n=m}^{\infty} \frac{1}{m!} y^{n-m} E\left[U.w_n(\xi_1' X).\left(((\xi_* - \xi_0)' X\right)^n\right] \Bigg|_{y \downarrow 0} \\ & = \frac{1}{m!} E\left[U.w_m(\xi_1' X).\left(((\xi_* - \xi_0)' X\right)^m\right]. \end{aligned}$$

[5] Using $(\mathrm{d}/\mathrm{d}y)^m y^n = \begin{cases} (n!/m!) y^{n-m} & \text{for } n \geq m, \\ 0 & \text{for } n < m. \end{cases}$

Consequently, for an arbitrarily small $\beta > 0$ and all nonnegative integers n,

$$E\left[U.w_n(\xi_1'X).\left(-\beta.(\xi_* - \xi_0)'X\right)^n\right] = 0$$

as well, so that

$$\begin{aligned}
0 &= \sum_{n=0}^{\infty} \frac{1}{n!} E\left[U.w_n(\xi_1'X).\left(-\beta.(\xi_* - \xi_0)'X\right)^n\right] \\
&= E\left[U.\sum_{n=0}^{\infty} \frac{1}{n!} w_n(\xi_1'X).\left(-\beta.(\xi_* - \xi_0)'X\right)^n\right] \\
&= E\left[U.w(\xi_1'X - \beta(\xi_* - \xi_0)'X)\right] \\
&= E\left[U.w((\xi_*'X + (\lambda_1 - \beta)(\xi_* - \xi_0))'X)\right] \\
&= \rho(\lambda_1 - \beta|\xi_*),
\end{aligned}$$

where the fourth equality follows by substituting $\xi_1 = \xi_* + \lambda_1(\xi_* - \xi_0)$. However, $\rho(\lambda_1 - \beta|\xi_*) \neq 0$ because λ_1 was the smallest λ for which $\rho(\lambda|\xi_*) = 0$. This contradiction implies that for some small $\varepsilon > 0$, $\rho(\lambda|\xi_*) \neq 0$ for all $\lambda \in (\lambda_1 - \varepsilon, \lambda_1) \cup (\lambda_1, \lambda_1 + \varepsilon)$. The same applies to $\lambda_2 = \inf_{\lambda \in \Lambda(\xi_*):\lambda > \lambda_1} \lambda$, etcetera. Thus, the set $\Lambda(\xi_*)$ is countable and nowhere dense.

Summarizing, the following lemma has been proved.

Lemma 2.1. *Let $\xi_0 \in \mathbb{R}^k$ be such that $E[U.w(\xi_0'X)] \neq 0$, and let $\xi_* \in \mathbb{B}^k$ be arbitrary, where $\mathbb{B}^k$ is defined by* (2.8). *Under the conditions of Theorem 2.1 and Assumption 2.6 the set*

$$\Lambda(\xi_*) = \{\lambda > 0 : E[U.w((\xi_0 + \lambda(\xi_* - \xi_0))'X)] = 0\}$$

is either empty or countable and nowhere dense.

Recall that in the case $k = 1$ the set $\mathbb{B}^k$ consists of two points: $\mathbb{B}^1 = \{\xi_0 - 1, \xi_0 + 1\}$. In the case $k \geq 2$ the set $\mathbb{B}^k$ is $(k-1)$-dimensional, so that any point $\xi_* \in \mathbb{B}^k$ is a function of $k - 1$ variables. For example, in the case $k = 2$ let

$$\xi_*(v) = \xi_0 + (\sin(2\pi v), \cos(2\pi v))', \ v \in [0, 1],$$

and in the case $k = 3$, let

$$\xi_*(v) = \xi_0 + \begin{pmatrix} \sin(2\pi v_1)\sin(2\pi v_2) \\ \cos(2\pi v_1)\sin(2\pi v_2) \\ \cos(2\pi v_2) \end{pmatrix}, \ v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in [0, 1] \times [0, 1].$$

In the general case $k \geq 2$, let

$$\xi_*(v) = \xi_0 + \eta_k(v_1, ..., v_{k-1}),\ v = (v_1, ..., v_{k-1})' \in [0,1]^{k-1},$$

where the mapping

$$\eta_k(v_1, ..., v_{k-1}) : [0,1]^{k-1} \to \{x \in \mathbb{R}^k : ||x|| = 1\} \tag{2.9}$$

is constructed recursively by the scheme

$$\eta_m(v_1, ..., v_{m-1}) = \begin{pmatrix} \sin(2\pi v_{m-1}).\eta_{m-1}(v_1, ..., v_{m-2}) \\ \cos(2\pi v_{m-1}) \end{pmatrix}$$

for $m = 2, 3, ..., k$, starting from $\eta_2(v_1) = (\sin(2\pi v_1), \cos(2\pi v_1))'$.

Is is easy to verify that the mapping (2.9) is one-to-one, and thus so is $\xi_*(v)$.

To prove that the set $S = \{\xi \in \mathbb{R}^k : E[U.w(\xi' X)] = 0\}$ has Lebesgue measure zero, denote $g(\xi) = E[U.w(\xi' X)]$, and let $\widetilde{\xi} = \xi_0 + \widetilde{\lambda}(\xi_*(\widetilde{V}) - \xi_0)$, where $\widetilde{V}$ is a random drawing from the uniform distribution on $[0,1]^{k-1}$, and $\widetilde{\lambda}$ is a random drawing from the standard exponential distribution, for example. Then

$$\begin{aligned}
\Pr\left[g(\widetilde{\xi}) = 0\right] &= E\left[I\left(g(\xi_0 + \widetilde{\lambda}(\xi_*(\widetilde{V}) - \xi_0)) = 0\right)\right] \\
&= E\left(E\left[I\left(g(\xi_0 + \widetilde{\lambda}(\xi_*(\widetilde{V}) - \xi_0)) = 0\right)\middle| \widetilde{V}\right]\right) \\
&= E\left(\Pr[g(\xi_0 + \widetilde{\lambda}(\xi_*(\widetilde{V}) - \xi_0)) = 0|\widetilde{V}]\right) \\
&= E\left(\Pr[\widetilde{\lambda} \in \Lambda(\xi_*(\widetilde{V}))|\widetilde{V}]\right) \\
&= \Pr[\widetilde{\lambda} \in \Lambda(\xi_*(\widetilde{V}))] = 0,
\end{aligned}$$

because by Lemma 2.1, the set $\Lambda(\xi_*(\widetilde{V}))$ has zero Lebesgue measure and is nowhere dense. Consequently, $g(\xi) \neq 0$ a.e. on $\mathbb{R}^k$.

2.4. Comparison with Stinchcombe and White

The result of Theorem 2.3 in Stinchcombe and White (1998) can be translated in my notation as follows:

Theorem 2.3. *Under Assumptions 2.1, 2.2 and 2.6 the set*

$$S = \{(\varsigma, \xi')' \in \mathbb{R} \times \mathbb{R}^k : E[U.w(\varsigma + \xi' X)] = 0\}$$

has Lebesgue measure zero and is nowhere dense in $\mathbb{R}^{k+1}$.

The main difference between this result and Theorem 2.2 is that Stinchcombe and White only require condition (2.3) instead of the stronger non-polynomial condition (2.1) in my case. Apparently, the presence of the intercept ς is the crux why condition (2.3) is sufficient in Theorem 2.3.

Observe from the proof of Theorem 2.2 that condition (2.1) is not directly used, but only indirectly via the condition that there exists a $\xi_0 \in \mathbb{R}^k$ such that $E[U.w(\xi_0'X)] \neq 0$, which follows from Theorem 2.1.

As to the role of ς, observe from (2.2) that there exists an $\varepsilon > 0$, possibly depending on ς, such that for all $\xi \in N_0(\varepsilon)$ (with the latter defined in Theorem 2.1),

$$w(\varsigma + \xi'X) = \sum_{n=0}^{\infty} \frac{w_n(\varsigma)}{n!}(\xi'X)^n.$$

Thus if

$$\text{for some } \varsigma_0 \in \mathbb{R} \text{ the set } \{n \in \mathbb{N} : w_n(\varsigma_0) = 0\} \text{ is finite or empty,} \tag{2.10}$$

(c.f. Assumption 2.3) and if for an arbitrary small $\varepsilon > 0$,

$$\sup_{|u|<\varepsilon} |w_n(\varsigma_0 + u)| < \infty \text{ for all } n \in \mathbb{N}, \tag{2.11}$$

(c.f. Assumption 2.4), then it follows from Theorem 2.1 that there exists a $\xi_0 \in \mathbb{R}^k$ such that $E[U.w(\varsigma_0 + \xi_0'X)] \neq 0$, so that then Theorem 2.3 follows similar to the proof of Theorem 2.2, with $w(u)$ replaced by $w(\varsigma_0 + u)$.

Condition (2.11) follows from the continuity of $w_n(u)$ on $\mathbb{R}$, so we only need to focus on condition (2.10), as follows. Suppose that for some $m \in \mathbb{N}$, $w_m(0) = 0$. Then for all $u \neq 0$ in an arbitrarily small neighborhood of zero,

$$w(u) = \sum_{n=0}^{m-1} \frac{w_n(0)}{n!}u^n + \sum_{n=m+1}^{\infty} \frac{w_n(0)}{n!}u^n$$

and thus

$$\begin{aligned} w_m(u) &= \sum_{n=0}^{m-1} \frac{w_n(0)}{n!}(\mathrm{d}/\mathrm{du})^m u^n + \sum_{n=m+1}^{\infty} \frac{w_n(0)}{n!}(\mathrm{d}/\mathrm{du})^m u^n \\ &= \sum_{n=m+1}^{\infty} \frac{w_n(0)}{n!}u^{n-m} = u\sum_{n=0}^{\infty} \frac{w_{n+m+1}(0)}{(n+m+1)!}u^n. \end{aligned}$$

Suppose that for some arbitrarily small $\varepsilon > 0$, $w_m(u) \equiv 0$ on $(-\varepsilon, \varepsilon)$. Then for all nonnegative integers s,

$$\begin{aligned} 0 = (\mathrm{d}/\mathrm{d}u)^s w_m(u)/u &= (\mathrm{d}/\mathrm{d}u)^s \sum_{n=0}^{\infty} \frac{w_{n+m+1}(0)}{(n+m+1)!} u^n \bigg|_{u=0} \\ &= \sum_{n=0}^{\infty} \frac{w_{n+m+1}(0)}{(n+m+1)!} (\mathrm{d}/\mathrm{d}u)^s u^n \bigg|_{u=0} \\ &= \sum_{n=s}^{\infty} \frac{w_{n+m+1}(0)}{(n+m+1)!} \frac{n!}{s!} u^{n-s} \bigg|_{u=0} \\ &= \frac{w_{s+m+1}(0)}{(s+m+1)!}, \end{aligned}$$

hence $w_{m+1+s}(0) = 0$ for $s = 0, 1, 2, \ldots..$ But this implies that $w(u)$ is a polynomial on $(-\varepsilon, \varepsilon)$, which in its turn implies that $w(u)$ is a polynomial, as is not hard to verify.

Thus, for each u_0 for which $w_n(u_0) = 0$ there exists an u_1 arbitrarily close to u_0 such that $w_n(u_1) \neq 0$. This implies, by the continuity of $w_n(u)$, via a similar argument as in the proof of Theorem 2.2, that for each n the set $\mathcal{W}_n = \{u \in \mathbb{R} : w_n(u) = 0\}$ has Lebesgue measure zero and is nowhere dense. The same applies to the union $\cup_{n=1}^{\infty} \mathcal{W}_n$, because for a random drawing Z from a normal distribution, for example, $\Pr[Z \in \cup_{n=1}^{\infty} \mathcal{W}_n] \leq \sum_{n=1}^{\infty} \Pr[Z \in \mathcal{W}_n] = 0$. Consequently,

Lemma 2.2. *If $w(u)$ is non-polynomial and analytical on $\mathbb{R}$ then the set $\{\varsigma \in \mathbb{R} : w_n(\varsigma) = 0$ for some $n \in \mathbb{N}\}$ has Lebesgue measure zero and is nowhere dense,*

which implies condition (2.10). This completes the proof of Theorem 2.3.

Remark 2.3. Note that Stinchcombe and White (1998) derive further generalizations of their Theorem 2.3, with applications to consistent model specification testing and artificial neural net theory,[6] and that their proofs are based on advanced functional analysis, in particular Banach space theory.

[6] See Kuan and White (1994) and Kuan (2008) for reviews of ANN models, and Bierens (1994) for comments on Kuan and White (1994).

3. Preliminary Hilbert space results

The proofs of the Hilbert-Schmidt and Mercer in the next sections employ rather elementary Hilbert space theoretical results, which I will summarize here. But first let me introduce the two key Hilbert spaces involved.

Definition 3.1 *Given a probability measure μ on a Borel-subset Ξ of a Euclidean space, $L^2(\mu)$ denotes the Hilbert space of Borel measurable real functions $f(\xi)$ on Ξ satisfying $\int f(\xi)^2 \mathrm{d}\mu(\xi) < \infty$, endowed with the innerproduct $\langle f_1, f_2 \rangle = \int f_1(\xi) f_2(\xi) \mathrm{d}\mu(\xi)$ and associated norm $||f|| = \sqrt{\langle f, f \rangle}$ and metric $||f_1 - f_2||$.*

Definition 3.2 *With μ and Ξ as before, $L^2(\mu \times \mu)$ denotes the Hilbert space of Borel measurable real functions $f(\xi_1, \xi_2)$ on $\Xi \times \Xi$ satisfying $\int \int f(\xi_1, \xi_2)^2 \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2) < \infty$, endowed with the innerproduct $\langle f_1, f_2 \rangle = \int f_1(\xi_1, \xi_2) f_2(\xi_1, \xi_2) \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2)$ and associated norm $||f|| = \sqrt{\langle f, f \rangle}$ and metric $||f_1 - f_2||$.*

In these two cases I have used the same notation $\langle ., . \rangle$ for the inner-product, as it will be clear from the context which one applies.

Recall that, in general, a Hilbert space $\mathcal{H}$ is a vector space endowed with an innerproduct and associated norm and metric, such that every Cauchy sequence in $\mathcal{H}$ takes a limit in $\mathcal{H}$. Thus, calling the spaces $L^2(\mu)$ and $L^2(\mu \times \mu)$ Hilbert spaces implicitly means that the Cauchy sequence requirement holds for them, which is a standard Hilbert space result. See for example Bierens (2014).

The proofs the of Hilbert-Schmidt and Mercer theorems employ the following properties of the Hilbert spaces $L^2(\mu)$ and $L^2(\mu \times \mu)$.

Lemma 3.1. *There exists an orthonormal sequence $\{\eta_j\}_{j=1}^{\infty}$ in $L^2(\mu)$ such that for every $f \in L^2(\mu)$,*
(a) $\lim_{n \to \infty} ||f - f_n||$, *where* $f_n(\xi) = \sum_{j=1}^{n} \langle f, \eta_j \rangle \eta_j(\xi)$;
(b) $\sum_{j=1}^{\infty} (\langle f, \eta_j \rangle)^2 = ||f||^2 < \infty$;
(c) there exists a Borel subset Ξ_0 of Ξ with $\mu(\Xi_0) = 0$ such that pointwise in $\xi \in \Xi \backslash \Xi_0$, $f(\xi) = \lim_{n \to \infty} f_n(\xi) = \sum_{j=1}^{\infty} \langle f, \eta_j \rangle \eta_j(\xi)$.

Such an orthonormal sequence $\{\eta_j\}_{j=1}^{\infty}$ is called *complete* in $L^2(\mu)$, also called an *orthonormal base* of $L^2(\mu)$, and $L^2(\mu)$ is said to be spanned by $\{\eta_j\}_{j=1}^{\infty}$, denoted by

$$L^2(\mu) = \mathrm{span}(\{\eta_j\}_{j=1}^{\infty}).$$

Note that similar to orthonormal bases of Euclidean spaces, $\{\eta_j\}_{j=1}^{\infty}$ is not unique.

The results (a) and (b) are standard Hilbert space results and can be found in any textbook on Hilbert spaces, for example Young (1988). In particular, property (b) is a direct consequence of property (a), because $||f - f_n||^2 = ||f||^2 - \sum_{j=1}^{n}(\langle f, \eta_j \rangle)^2$.

Property (c) has been proved in Bierens (2014, Theorem 9). Its interpretation is that, with $\widetilde{\xi}$ a random drawing from the distribution μ, property (a) implies that $f(\widetilde{\xi}) = \lim_{n\to\infty} f_n(\widetilde{\xi})$ a.s. This motivates to write the result (c) as

$$f(\xi) = \sum_{j=1}^{\infty} \langle f, \eta_j \rangle \eta_j(\xi) \text{ a.s. } \mu. \tag{3.1}$$

Lemma 3.2. *Let $\{\eta_j\}_{j=1}^{\infty}$ be the orthonormal base of $L^2(\mu)$ in Lemma 3.1, and denote $\eta_{i,j}(\xi_1, \xi_2) = \eta_i(\xi_1)\eta_j(\xi_2)$. Then $\{\eta_{i,j}\}_{i,j=1}^{\infty}$ is an orthonormal base of $L^2(\mu \times \mu)$. Consequently, for every $g \in L^2(\mu \times \mu)$,*
(a) $\lim_{\min(n_1,n_2)\to\infty} \int \left(g(\xi_1, \xi_2) - \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{i,j}(g)\eta_i(\xi_1)\eta_j(\xi_2)\right)^2 \mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = 0$, *where* $c_{i,j}(g) = \langle g, \eta_{i,j} \rangle = \int\int \eta_i(\xi_1)g(\xi_1, \xi_2)\eta_j(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$;
(b) $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(c_{i,j}(g))^2 = ||g||^2 = \int\int g(\xi_1, \xi_2)^2 \mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) < \infty$;
(c) $g(\xi_1, \xi_2) = \lim_{\min(n_1,n_2)\to\infty} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{i,j}(g)\eta_i(\xi_1)\eta_j(\xi_2)$ *pointwise in* $(\xi_1, \xi_2) \in \Xi \times \Xi \backslash \Xi_0 \times \Xi_0$, *where Ξ_0 is the same as in Lemma 3.1.*

The claim that the double sequence $\{\eta_i(\xi_1)\eta_j(\xi_2)\}_{i,j=1}^{\infty}$ is complete in $L^2(\mu \times \mu)$ and thus

$$L^2(\mu \times \mu) = \mathrm{span}(\{\eta_i(\xi_1)\eta_j(\xi_2)\}_{i,j=1}^{\infty}),$$

is also a standard Hilbert space result, but can easily be proved similar to Theorem 17 in Bierens (2014). Again, the interpretation of result (c) is that, with $\widetilde{\xi}_1$ and $\widetilde{\xi}_2$ independent random drawings from the distribution of μ,

$$g(\widetilde{\xi}_1, \widetilde{\xi}_2) = \lim_{\min(n_1,n_2)\to\infty} \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} c_{i,j}(g)\eta_i(\widetilde{\xi}_1)\eta_j(\widetilde{\xi}_2) \text{ a.s.},$$

which motivates to write result (c) as

$$g(\xi_1, \xi_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} c_{i,j}(g)\eta_i(\xi_1)\eta_j(\xi_2) \text{ a.s. } \mu \times \mu. \tag{3.2}$$

However, from now on I will no longer mention the caveats "a.s. μ" and "a.s. $\mu \times \mu$" in equalities of the type (3.1) and (3.2), respectively, but instead only mention "for all $\xi \in \Xi$" or "for all $(\xi_1, \xi_2) \in \Xi \times \Xi$" if these caveats do not apply.

Also, I will need the well-known projection theorem.

Lemma 3.3. (Projection theorem) *Let $\mathcal{S}$ be a sub-Hilbert space of a Hilbert space $\mathcal{H}$, and let $y \in \mathcal{H}$ be arbitrary. Let x be the projection of y on $\mathcal{S}$, defined as an $x \in \mathcal{H}$ such that $||y - x|| = \inf_{z \in \mathcal{S}} ||y - z||$. Then x is unique and is contained in $\mathcal{S}$. Moreover, the projection residual $u = y - x$ is orthogonal to $\mathcal{S}$, in the sense that for all $z \in \mathcal{S}$, $\langle x, z \rangle = 0$.*

Proof. See for example Bierens (2004, Theorem 7.A.3, p.202) or Bierens (2014, Theorem 6). However, the projection theorem is one of the most important and useful Hilbert space theoretical results, and its proof can be found in most text-books on Hilbert spaces.[7] ■

In particular, if $\mathcal{S}$ is of the form $\mathcal{S} = \text{span}(\{z_m\}_{m=1}^{\infty})$, where $\{z_m\}_{m=1}^{\infty}$ is a possibly incomplete orthonormal sequence in $\mathcal{H}$, then the projection x of $y \in \mathcal{H}$ on $\mathcal{S}$ takes the form $x = \sum_{m=1}^{\infty} \langle y, z_m \rangle z_m$, in the sense that $\lim_{n \to \infty} \|x - \sum_{m=1}^{n} \langle y, z_m \rangle z_m\| = 0$.

Finally, the interpretation of $\Gamma(\xi_1, \xi_2)$ as a covariance function of a zero-mean Gaussian process is not essential for the Mercer and Hilbert-Schmidt theorems. All that matters is the following.

Assumption 3.1. *Let Ξ be a compact subset of a Euclidean space. The function $\Gamma(\xi_1, \xi_2)$ on $\Xi \times \Xi$ is real-valued, continuous, symmetric, and positive semidefinite with respect to a probability measure μ on Ξ.*

The latter means that

$$\int \int \phi(\xi_1) \Gamma(\xi_1, \xi_2) \phi(\xi_2) \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2) \geq 0 \text{ for all } \phi \in L^2(\mu). \tag{3.3}$$

Of course, symmetry means that $\Gamma(\xi_1, \xi_2) = \Gamma(\xi_2, \xi_1)$.

[7] Surprisingly, the book by Young (1988) is one of the exceptions.

Remark 3.1. An alternative definition of symmetric positive semidefinite kernels $\Gamma(\xi_1, \xi_2)$ on $\Xi \times \Xi$ is that for each $n \in \mathbb{N}$ and arbitrary $\xi_1, \xi_2, ..., \xi_n$ in Ξ the $n \times n$ matrix with elements $\Gamma(\xi_i, \xi_j)$, $i, j = 1, 2, .., n$, is symmetric positive semidefinite, together with the conditions that Ξ is compact and $\Gamma(\xi_1, \xi_2)$ is continuous and real valued on $\Xi \times \Xi$. To prove that this alternative definition implies (3.3), let $\widetilde{\xi}_i$, $i = 1, 2, .., n$, be independent random drawings from the distribution μ. Then for $\phi \in L^2(\mu)$,

$$\int\int \phi(\xi_1)\Gamma(\xi_1, \xi_2)\phi(\xi_2)\,\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = p\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi\left(\widetilde{\xi}_i\right)\Gamma(\widetilde{\xi}_i, \widetilde{\xi}_j)\phi\left(\widetilde{\xi}_j\right),$$

as is not hard (but rather tedious) to verify. Since by this alternative definition, $\sum_{i=1}^{n}\sum_{j=1}^{n}\phi\left(\widetilde{\xi}_i\right)\Gamma(\widetilde{\xi}_i, \widetilde{\xi}_j)\phi\left(\widetilde{\xi}_j\right) \geq 0$ a.s., (3.3) follows.

4. The Hilbert-Schmidt theorem

The existence of eigenfunctions and corresponding eigenvalues of the function $\Gamma(\xi_1, \xi_2)$ in Assumption 3.1 was proved by David Hilbert and his Ph.D. student Erhard Schmidt in the period 1904-1908 for the case $\Xi = [a, b]$ and μ the Lebesgue measure. Rather than reproducing their proof, which is lengthy and complicated, I will in this section provide an alternative proof of the following version of the Hilbert-Schmidt theorem.

Theorem 4.1. (Hilbert-Schmidt theorem) *Let the conditions on* $\Gamma(\xi_1, \xi_2)$, Ξ *and* μ *in Assumption 3.1 be satisfied. Consider the following eigenvalue problem.*

$$\begin{aligned}&\textit{Find a scalar } \lambda \textit{ and a function } \psi \in L^2(\mu) \textit{ with } ||\psi|| = 1 \textit{ such that}\\ &\lambda\psi(\xi_1) = \int \Gamma(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_2) \textit{ for all } \xi_1 \in \Xi.\end{aligned} \tag{4.1}$$

Then the following hold:
(*a*) *This eigenvalue problem has countable many solution* $\{(\lambda_m, \psi_m)\}_{m=1}^{\infty}$;
(*b*) *The eigenvalues* λ_m *are real-valued and nonnegative;*
(*c*) *The eigenfunctions* ψ_m *corresponding to positive eigenvalues* λ_m *are continuous on* Ξ;[8]

[8] The claim in Lemma 1 in BP that *all* the eigenfunctions are continuous is not correct!

(*d*) *The sequence of eigenfunctions* $\{\psi_m\}_{m=1}^{\infty}$, *including the eigenfunctions corresponding to zero eigenvalues, is orthonormal.*
(*e*) *If all the eigenvalues are zero then* $\Gamma(\xi_1, \xi_2) \equiv 0$ *on* $\Xi \times \Xi$.

Note that, given part (a), the parts (b) and (c) are nearly trivial because for any solution (λ, ψ) of (4.1),

$$\lambda = \lambda \int \psi(\xi_1)^2 \mathrm{d}\mu(\xi_1) = \int \int \psi(\xi_1)\Gamma(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \geq 0$$

and if $\lambda > 0$ then $\psi(\xi_1) = \lambda^{-1} \int \Gamma(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_2)$ is continuous on Ξ because $\Gamma(\xi_1, \xi_2)$ is continuous on $\Xi \times \Xi$.

As to part (d), let λ_1, λ_2 be a pair of eigenvalues with corresponding eigenfunctions ψ_1, ψ_2. Then

$$\begin{aligned} \lambda_1 \int \psi_1(\xi_1)\psi_2(\xi_1)\mathrm{d}\mu(\xi_1) &= \int \psi_2(\xi_1) \int \Gamma(\xi_1, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_2)\mathrm{d}\mu(\xi_1) \\ &= \int \left(\int \psi_2(\xi_1)\Gamma(\xi_1, \xi_2)\mathrm{d}\mu(\xi_1) \right) \psi_1(\xi_2)\mathrm{d}\mu(\xi_2) \\ &= \lambda_2 \int \psi_2(\xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_2), \end{aligned}$$

hence if $\lambda_1 \neq \lambda_2$ then $\langle \psi_1, \psi_2 \rangle = \int \psi_1(\xi_1)\psi_2(\xi_1)\mathrm{d}\mu(\xi_1) = 0$. Thus, the eigenfunctions corresponding to distinct eigenvalues are orthogonal.

Now consider the case $\lambda_1 = \lambda_2 = \lambda \geq 0$, with distinct eigenfunctions ψ_1 and ψ_2. Then $\psi_1^* = \psi_1 - ||\psi_2||^{-2} \langle \psi_1, \psi_2 \rangle \psi_2$ is also an eigenfunction with eigenvalue λ, and $\langle \psi_1^*, \psi_2 \rangle = 0$. The extension of this argument to more than two common eigenvalues is not too hard. Thus, the eigenfunctions of $\Gamma(\xi_1, \xi_2)$ are (or can be chosen) orthonormal.

4.1. The positive eigenvalue problem

In first instance I will focus on the *positive* eigenvalue problem:

$$\textit{Find a } \lambda > 0 \textit{ and a function } \psi \in L^2(\mu) \textit{ with } ||\psi|| = 1 \textit{ such that}$$
$$\lambda\psi(\xi_1) = \int \Gamma(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_2) \textit{ for all } \xi_1 \in \Xi. \tag{4.2}$$

Denote

$$G(\psi, \lambda) = \int \left(\int \Gamma(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_2) - \lambda\psi(\xi_1) \right)^2 \mathrm{d}\mu(\xi_1).$$

Then the eigenvalue problem (4.2) is equivalent to the problem:

$$\textit{Find a } \lambda > 0 \textit{ and a } \psi \in L^2(\mu),\ ||\psi|| = 1,\ \textit{such that } G(\psi, \lambda) = 0. \tag{4.3}$$

Remark 4.1. Strictly speaking, a solution (λ, ψ) of problem (4.3) implies that (4.1) holds for all ξ_1 in the support $S(\mu)$ of the distribution μ. In other words, denoting $f(\xi|\psi, \lambda) = \int \Gamma(\xi, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_2) - \lambda\psi(\xi)$, with (ψ, λ) a solution of (4.3), the latter implies that $f(\xi|\psi, \lambda) = 0$ for all $\xi \in S(\mu)$. Now write $\psi_1(\xi) = \psi(\xi)I(\xi \in S(\mu))$, $\psi_2(\xi) = \psi(\xi)I(\xi \in \Xi \backslash S(\mu))$, where $I(.)$ is the well-known indicator function, and note that

$$\int \Gamma(\xi, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_2) \equiv \int \Gamma(\xi, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_2) \text{ on } \Xi.$$

Thus, the solution of (4.3) implies that $I(\xi \in S(\mu)) \int \Gamma(\xi, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_2) = \lambda\psi_1(\xi)$ on Ξ, so that

$$f(\xi|\psi, \lambda) = I(\xi \in \Xi \backslash S(\mu)) \int \Gamma(\xi, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_2) - \lambda\psi_2(\xi).$$

Since ψ_1 and λ are determined by (4.3), we may without loss of generality *choose*

$$\psi_2(\xi) = \lambda^{-1}I(\xi \in \Xi \backslash S(\mu)) \int \Gamma(\xi, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_2)$$

and set $\psi(\xi) = \psi_1(\xi) + \psi_2(\xi)$, so that then $f(\xi|\psi, \lambda) = 0$ for all $\xi \in \Xi$. Finally, note that $\psi_1(\xi).\psi_2(\xi) = 0$ and $\int \psi_1(\xi)^2\mathrm{d}\mu(\xi) = 1$, $\int \psi_2(\xi)^2\mathrm{d}\mu(\xi) = 0$, hence $||\psi|| = ||\psi_1 + \psi_2|| = ||\psi_1|| = 1$.

Thus, indeed, the eigenvalue problems (4.2) and (4.3) are equivalent regardless the nature of the probability measure μ.

Given ψ, the function $G(\psi, \lambda)$ is quadratic in λ:

$$\begin{aligned} G(\psi, \lambda) &= \int\int \psi(\xi_1)\Gamma_*(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\ &\quad -2\lambda \int\int \psi(\xi_1)\Gamma(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) + \lambda^2, \text{ where} \\ \Gamma_*(\xi_1, \xi_2) &= \int \Gamma(\xi_1, \xi)\Gamma(\xi, \xi_2)\mathrm{d}\mu(\xi), \end{aligned}$$

which is minimal for

$$\lambda = \int\int \psi(\xi_1)\Gamma(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2). \tag{4.4}$$

Substituting this solution in $G(\psi,\lambda)$ yields

$$\begin{aligned}\underline{G}(\psi) &= \inf_{\lambda} G(\psi,\lambda) = \int\int \psi(\xi_1)\Gamma_*(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)\\ &\quad - \left(\int\int \psi(\xi_1)\Gamma(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)\right)^2.\end{aligned}$$

Thus, problem (4.3) now becomes:

$$\begin{aligned}&\textit{Find a } \psi \in L^2(\mu),\ ||\psi|| = 1,\ \textit{such that } \underline{G}(\psi) = 0\\ &\textit{and then determine } \lambda \textit{ by } (4.4).\end{aligned} \tag{4.5}$$

4.2. The maximum eigenvalue problem

Equation (4.4) and problem (4.5) suggest that the ψ for which (4.4) is maximal may be a suitable candidate for the eigenfunction corresponding to the largest eigenvalues, because a similar trick works for symmetric positive semidefinite matrices. Thus, the idea is to use $\psi_1 = \arg\max_{\psi\in L^2(\mu),\ ||\psi||=1} \int\int \psi(\xi_1)\Gamma(\xi_1,\xi_2)\psi(\xi_2)$ $\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$ as a possible candidate for the eigenfunction corresponding to the largest eigenvalue, and then check whether $\underline{G}(\psi_1) = 0$. Indeed, the latter holds, as will be shown now.

As in Lemma 3.1, let $\{\eta_m\}_{m=1}^{\infty}$ be an orthonormal base of $L^2(\mu)$. Then by Lemma 3.2, $\Gamma(\xi_1,\xi_2)$ has the series representation

$$\Gamma(\xi_1,\xi_2) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \alpha_{i,j}\eta_i(\xi_1)\eta_j(\xi_2), \tag{4.6}$$

where

$$\alpha_{i,j} = \int\int \eta_i(\xi_1)\Gamma(\xi_1,\xi_2)\eta_j(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$$

with $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\alpha_{i,j}^2 < \infty$. It is easy to verify that (4.6) implies

$$\Gamma_*(\xi_1,\xi_2) = \int \Gamma(\xi_1,\xi)\Gamma(\xi,\xi_2)\mathrm{d}\mu(\xi) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left(\sum_{m=1}^{\infty}\alpha_{i,m}\alpha_{m,j}\right)\eta_i(\xi_1)\eta_j(\xi_2). \tag{4.7}$$

Denote

$$\Gamma_n(\xi_1,\xi_2) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{i,j}\eta_i(\xi_1)\eta_j(\xi_2),\ \ \eta^{(n)}(\xi) = (\eta_1(\xi),\eta_2(\xi),...,\eta_n(\xi))',$$

and let A_n be the $n\times n$ matrix with elements $\alpha_{i,j}$, $i,j=1,2,..,n$. Then

$$\Gamma_n(\xi_1,\xi_2) = \eta^{(n)}(\xi)'A_n\eta^{(n)}(\xi).$$

Since $\Gamma(\xi_1,\xi_2)$ is symmetric positive semidefinite, so is A_n for any $n\in\mathbb{N}$.

As is well-known, the maximum eigenvalue of A_n is

$$\lambda_{1,n} = \sup_{x\in\mathbb{R}^n:||x||=1} x'A_nx, \tag{4.8}$$

with corresponding normalized eigenvector

$$q_n = (q_{1,n},...,q_{n,n})' = \arg\max_{x\in\mathbb{R}^n:||x||=1} x'A_nx, \tag{4.9}$$

Therefore, $\lambda_{1,n}$ is the maximum eigenvalue of $\Gamma_n(\xi_1,\xi_2)$, with corresponding eigenfunction

$$\psi_{1,n}(\xi) = \sum_{i=1}^{n} q_{i,n}\eta_i(\xi) = q_n'\eta^{(n)}(\xi).$$

Then

$$\begin{aligned}
\int\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_2) &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\alpha_{i,j}\eta_i(\xi_1)\int\sum_{m=1}^{n}q_{m,n}.\eta_m(\xi_2)\eta_j(\xi_2)\mathrm{d}\mu(\xi_2)\\
&= \sum_{i=1}^{\infty}\sum_{j=1}^{n}\eta_i(\xi_1)\alpha_{i,j}q_{j,n}\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\eta_i(\xi_1)\alpha_{i,j}q_{j,n} + \sum_{i=n+1}^{\infty}\sum_{j=1}^{n}\eta_i(\xi_1)\alpha_{i,j}q_{j,n}\\
&= \eta^{(n)}(\xi)'A_nq_n + \sum_{i=n+1}^{\infty}\sum_{j=1}^{n}\eta_i(\xi_1)\alpha_{i,j}q_{j,n}\\
&= \lambda_{1,n}\eta^{(n)}(\xi)'q_n + \sum_{i=n+1}^{\infty}\sum_{j=1}^{n}\eta_i(\xi_1)\alpha_{i,j}q_{j,n}\\
&= \lambda_{1,n}\psi_{1,n}(\xi_1) + \rho_n(\xi_1|\psi_{1,n})
\end{aligned}$$

where

$$\rho_n(\xi|\psi_{1,n}) = \sum_{i=n+1}^{\infty}\left(\sum_{j=1}^{n}\alpha_{i,j}q_{j,n}\right)\eta_i(\xi).$$

Note that $\int \rho_n(\xi|\psi_{1,n})\psi_{1,n}(\xi)\mathrm{d}\mu(\xi) = 0$, hence

$$\lambda_{1,n} = q_n' A_n q_n = \int\int \psi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2),$$

which by (4.8) is monotonic nondecreasing in n, and bounded from above by $\sup_{(\xi_1,\xi_2)\in\Xi\times\Xi}\Gamma(\xi_1,\xi_2)$, so that $\lim_{n\to\infty}\lambda_{1,n} = \lambda_1$ exists and is finite, with

$$\begin{aligned}\lambda_1 &= \lim_{n\to\infty}\int\int \psi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)\\ &= \sup_{n\geq 1}\int\int \psi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)\\ &= \sup_{n\geq 1}\sup_{\psi\in\mathrm{span}(\{\eta_i\}_{i=1}^{n}),\ ||\psi||=1}\int\int \psi(\xi_1)\Gamma(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2). \quad (4.10)\end{aligned}$$

The latter follows from (4.9). Now (4.10) implies that

Lemma 4.1. $\lambda_1 = \sup_{\psi\in L^2(\mu).||\psi||=1}\int\int \psi(\xi_1)\Gamma(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$,

as will be proved below. This result suggests that

$$\psi_1 = \arg\max_{\psi\in L^2(\mu).||\psi||=1}\int\int \psi(\xi_1)\Gamma(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \qquad (4.11)$$

is a possible eigenfunction corresponding to eigenvalue λ_1. This is indeed the case because

Lemma 4.2. $\lambda_1^2 = \int\int \psi_1(\xi_1)\Gamma_*(\xi_1,\xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$,

as will be proved below as well, hence $G(\psi_1,\lambda_1) = 0$.

Summarizing, the following result holds.

Lemma 4.3. *Let the conditions on $\Gamma(\xi_1,\xi_2)$ in Assumption 3.1 be satisfied. Let ψ_1 be a solution of (4.11) and let $\lambda_1 = \int\int \psi_1(\xi_1)\Gamma(\xi_1,\xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$.*

If $\lambda_1 > 0$ then λ_1 is the maximum eigenvalue of $\Gamma(\xi_1, \xi_2)$ with corresponding eigenfunction ψ_1, whereas if $\lambda_1 = 0$ then $\Gamma(\xi_1, \xi_2) \equiv 0$.

The latter follows from the fact that if $\lambda_1 = 0$ then $A_n = O$ for all $n \in \mathbb{N}$, hence $\alpha_{i,j} = 0$ for all $i, j \in \mathbb{N}$. Therefore, pending the proofs of Lemmas 4.1 and 4.2, part (d) of Theorem 4.1 has been proved.

4.3. The other eigenvalues and eigenfunctions.

Given that $\lambda_1 > 0$, let

$$\Gamma^{(2)}(\xi_1, \xi_2) = \Gamma(\xi_1, \xi_2) - \lambda_1 \psi_1(\xi_1)\psi_1(\xi_2)$$

which is symmetric and continuous on $\Xi \times \Xi$. To prove that $\Gamma^{(2)}(\xi_1, \xi_2)$ is positive semidefinite, let $\phi \in L^2(\mu)$ be arbitrary. We can write $\phi = \langle \phi, \psi_1 \rangle \psi_1 + r$, where $\langle r, \psi_1 \rangle = 0$,[9] hence

$$\begin{aligned}
&\int\int \phi(\xi_1)\Gamma^{(2)}(\xi_1, \xi_2)\phi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&= \int\int \phi(\xi_1)\Gamma(\xi_1, \xi_2)\phi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) - \lambda_1 \left(\int \phi(\xi)\psi_1(\xi)\mathrm{d}\mu(\xi)\right)^2 \\
&= \langle \phi, \psi_1 \rangle^2 \int\int \psi_1(\xi_1)\Gamma(\xi_1, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&\quad - 2\langle \phi, \psi_1 \rangle \int\int r(\xi_1)\Gamma(\xi_1, \xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&\quad + \int\int r(\xi_1)\Gamma(\xi_1, \xi_2)r(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) - \lambda_1 \langle \phi, \psi_1 \rangle^2 \\
&= \int\int r(\xi_1)\Gamma(\xi_1, \xi_2)r(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \geq 0.
\end{aligned}$$

It follows now from Lemma 4.3 that the eigenfunction ψ_2 corresponding to the second largest eigenvalue λ_2 of $\Gamma(\xi_1, \xi_2)$ is a solution of

$$\psi_2 = \arg \max_{\psi \in L^2(\mu), ||\psi||=1} \int\int \psi(\xi_1)\Gamma^{(2)}(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2),$$

with eigenvalue

$$\lambda_2 = \int\int \psi_2(\xi_1)\Gamma^{(2)}(\xi_1, \xi_2)\psi_2(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \leq \lambda_1.$$

[9] Note that r is the residual of the projection of ϕ on $\mathrm{span}(\psi_1)$.

More generally, let

$$\Gamma^{(m+1)}(\xi_1, \xi_2) = \Gamma(\xi_1, \xi_2) - \sum_{i=1}^{m} \lambda_i \psi_i(\xi_1)\psi_i(\xi_2)$$

for $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_m > 0$. Similar to $\Gamma^{(2)}(\xi_1, \xi_2)$ it is easy to verify that $\Gamma^{(m+1)}(\xi_1, \xi_2)$ is symmetric positive semidefinite[10] and continuous on $\Xi \times \Xi$. Then by Lemma 4.3,

$$\psi_{m+1} = \arg \max_{\psi \in L^2(\mu), ||\psi||=1} \int \int \psi(\xi_1)\Gamma^{(m+1)}(\xi_1, \xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$$

is an eigenfunction of $\Gamma(\xi_1, \xi_2)$ with eigenvalue

$$\lambda_{m+1} = \int \int \psi_{m+1}(\xi_1)\Gamma^{(m+1)}(\xi_1, \xi_2)\psi_{m+1}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \leq \lambda_m. \tag{4.12}$$

If $\lambda_{m+1} > 0$, repeat this procedure for the next m. Otherwise, $\Gamma^{(m+1)}(\xi_1, \xi_2) \equiv 0$, hence $\Gamma(\xi_1, \xi_2) \equiv \sum_{i=1}^{m} \lambda_i \psi_i(\xi_1)\psi_i(\xi_2)$. In the latter case all the functions in the orthogonal complement $S_{1,m}^{\perp}$ of $S_{1,m} = \mathrm{span}(\{\psi_i\}_{i=1}^{m})$ are eigenfunctions of Γ with zero eigenvalues, and using the approach in Remark 5.1 below one can construct an orthonormal sequence $\{\psi_i\}_{i=m+1}^{\infty}$ such that $S_{1,m}^{\perp} = \mathrm{span}(\{\psi_i\}_{i=m+1}^{\infty})$. Clearly, $\{\psi_i\}_{i=m+1}^{\infty}$ is then a sequence of eigenfunction of Γ with zero eigenvalues, and thus $\{\psi_i\}_{i=1}^{\infty}$ is the countable infinite orthonormal sequence of all eigenfunctions.

In the case that we can repeat the procedure (4.12) indefinitely it yields a countable infinite orthonormal sequence $\{\psi_i\}_{i=1}^{\infty}$ of eigenfunctions with positive eigenvalues. However, this does not mean that there are no zero eigenvalues. Actually, if the complement $S_1^{\perp}$ of $S_1 = \mathrm{span}(\{\psi_i\}_{i=1}^{\infty})$ is nontrivial, $S_1^{\perp} \neq \{0\}$, then by Mercer's theorem below, $S_1^{\perp}$ is spanned by an orthonormal sequence of eigenfunctions corresponding to zero eigenvalues.

Pending the proofs of Lemmas 4.1 and 4.2 below, Theorem 4.1 has now been proved.

4.4. Proof of Lemma 4.1

Let ψ_1 be a solution of (4.11). Since $\psi_1 \in L^2(\mu)$ it has the series representation

$$\psi_1(\xi) = \sum_{m=1}^{\infty} c_m \eta_m(\xi), \tag{4.13}$$

[10] Project ϕ on $\mathrm{span}(\{\psi_i\}_{i=1}^{m})$. The projection is $\sum_{i=1}^{n} \langle \phi, \psi_i \rangle \psi_i$ with residual r. Then show that $\int \int \phi(\xi_1)\Gamma^{(m+1)}(\xi_1, \xi_2)\phi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = \int \int r(\xi_1)\Gamma(\xi_1, \xi_2)r(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$.

$$\text{where } c_m = \int \psi_1(\xi)\eta_m(\xi)\mathrm{d}\mu(\xi), \ \sum_{m=1}^{\infty} c_m^2 = 1.$$

Denote

$$\begin{aligned} \psi_1^{(n)}(\xi) &= \sum_{m=1}^{n} \gamma_{n,m}\eta_m(\xi), \ \text{where } \gamma_{n,m} = \frac{c_m}{\sqrt{\sum_{i=1}^n c_i^2}}, \\ \gamma_n &= (\gamma_{n,1}, \gamma_{n,2}, ..., \gamma_{n,n})' \end{aligned} \quad (4.14)$$

and note that

$$\begin{aligned} &\int\int \psi_1^{(n)}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_1^{(n)}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = \gamma_n' A_n \gamma_n \\ &\quad \le q_n' A_n q_n \\ &\quad\quad = \int\int \psi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \end{aligned}$$

and

$$\lim_{n\to\infty} \int \left(\psi_1(\xi) - \psi_1^{(n)}(\xi)\right)^2 \mathrm{d}\mu(\xi) = 0. \quad (4.15)$$

Hence

$$\begin{aligned} &\int\int \psi_1(\xi_1)\Gamma(\xi_1,\xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\ &= \limsup_{n\to\infty} \int\int \psi_1^{(n)}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_1^{(n)}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\ &\le \limsup_{n\to\infty} \int\int \psi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = \lambda_1 \end{aligned}$$

Since (4.10) also implies

$$\int\int \psi_1(\xi_1)\Gamma(\xi_1,\xi_2)\psi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \ge \lambda_1,$$

Lemma 4.1 follows.

4.5. Proof of Lemma 4.2

First note that

$$\int \rho_n(\xi|\psi_{1,n})^2\mathrm{d}\mu(\xi) = \sum_{i=n+1}^{\infty}\left(\sum_{j=1}^{n}\alpha_{i,j}q_{j,n}\right)^2 \le \sum_{m=n+1}^{\infty}\left(\sum_{i=1}^{n}\alpha_{i,m}^2\right) = o(1), \quad (4.16)$$

where the inequality follows from Schwartz inequality and $\sum_{j=1}^{n} q_{j,n}^2 = 1$,[11] and the last equality follows from $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\alpha_{i,j}^2 < \infty$.

Next, observe from (4.7) that

$$\begin{aligned}
&\int\int \psi_{1,n}(\xi_1)\Gamma_*(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{m=1}^{\infty}\alpha_{i,m}\alpha_{m,j}\right) q_{i,n}q_{j,n} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{m=1}^{n}\alpha_{i,m}\alpha_{m,j}\right) q_{i,n}q_{j,n} + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\sum_{m=n+1}^{\infty}\alpha_{i,m}\alpha_{m,j}\right) q_{i,n}q_{j,n} \\
&= q_n' A_n^2 q_n + \sum_{m=n+1}^{\infty}\left(\sum_{i=1}^{n}\alpha_{i,m}q_{i,n}\right)^2 \\
&= \lambda_{1,n} q_n' A_n q_n + \sum_{m=n+1}^{\infty}\left(\sum_{i=1}^{n}\alpha_{i,m}q_{i,n}\right)^2 \\
&= \lambda_{1,n}\int\int \psi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&\qquad\qquad + \int \rho_n(\xi|\psi_{1,n})^2\mathrm{d}\mu(\xi). \qquad (4.17)
\end{aligned}$$

Hence,

$$\lim_{n\to\infty}\int\int \psi_{1,n}(\xi_1)\Gamma_*(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = \lambda_1^2. \qquad (4.18)$$

Let

$$\phi_1 = \arg\max_{\psi\in L^2(\mu).||\psi||=1}\int\int \psi(\xi_1)\Gamma_*(\xi_1,\xi_2)\psi(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2),$$

[11] I.e.,

$$\begin{aligned}
\left|\sum_{j=1}^{n}\alpha_{i,j}q_{j,n}\right| &= n\left|\frac{1}{n}\sum_{j=1}^{n}\alpha_{i,j}q_{j,n}\right| \le n\sqrt{\frac{1}{n}\sum_{j=1}^{n}\alpha_{i,j}^2}\sqrt{\frac{1}{n}\sum_{j=1}^{n}q_{j,n}^2} \\
&= \sqrt{\sum_{j=1}^{n}\alpha_{i,j}^2}\sqrt{\sum_{j=1}^{n}q_{j,n}^2} = \sqrt{\sum_{j=1}^{n}\alpha_{i,j}^2}.
\end{aligned}$$

and suppose that $\phi_1 \neq \psi_1$, so that

$$\int\int \phi_1(\xi_1)\Gamma(\xi_1,\xi_2)\phi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) < \lambda_1. \tag{4.19}$$

Similar to (4.13) and (4.14) we can write

$$\begin{aligned}
\phi_1(\xi) &= \sum_{m=1}^{\infty} d_m\eta_m(\xi), \text{ where } d_m = \int \phi_1(\xi)\eta_m(\xi)\mathrm{d}\mu(\xi), \ \sum_{m=1}^{\infty} d_m^2 = 1, \\
\phi_{1,n}(\xi) &= \sum_{m=1}^{n} \delta_{n,m}\eta_m(\xi), \text{ where } \delta_{n,m} = \frac{d_m}{\sqrt{\sum_{i=1}^{n} d_i^2}}, \\
\delta_n &= (\delta_{n,1}, \delta_{n,2}, ..., \delta_{n,n})',
\end{aligned}$$

and similar to (4.15) we have

$$\lim_{n\to\infty} \int (\phi_1(\xi) - \phi_{1,n}(\xi))^2 \mathrm{d}\mu(\xi) = 0,$$

as is not hard to verify. It follows now from (4.16), (4.17), (4.18) and (4.19) that

$$\begin{aligned}
\lambda_1^2 &= \lim_{n\to\infty} \int\int \psi_{1,n}(\xi_1)\Gamma_*(\xi_1,\xi_2)\psi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&\geq \lim_{n\to\infty} \lambda_{1,n} \lim_{n\to\infty} \int\int \phi_{1,n}(\xi_1)\Gamma(\xi_1,\xi_2)\phi_{1,n}(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\
&\quad + \lim_{n\to\infty} \int \rho_n(\xi|\psi_{1,n})^2\mathrm{d}\mu(\xi) \\
&= \lambda_1 \int\int \phi_1(\xi_1)\Gamma(\xi_1,\xi_2)\phi_1(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) > \lambda_1^2,
\end{aligned}$$

which proves that (4.19) is not possible. Consequently, $\phi_1 = \psi_1$, which proves Lemma 4.2.

5. Mercer's theorem

As said before, it was claimed in BP that Lemma 1 is a straightforward further elaboration of Mercer's theorem, with a reference to Dunford and Schwartz (1963, p. 1088)), mimicking the properties of eigenvalues and eigenvectors of positive definite symmetric matrices. However, the latter reference refers to an exercise.

Therefore, in this section I will provide the proof of Mercer's theorem, restated here as follows.

Theorem 5.1. *Let the conditions on $\Gamma(\xi_1, \xi_2)$, Ξ and μ in Assumption 3.1 be satisfied. Let $\{\psi_m\}_{m=1}^{\infty}$ be the sequence of orthonormal solutions of the eigenvalue problem*

$$\lambda_m \psi_m(\xi_1) = \int \Gamma(\xi_1, \xi_2)\psi_m(\xi_2)\mathrm{d}\mu(\xi_2) \text{ for all } \xi_1 \in \Xi, \tag{5.1}$$

with corresponding sequence $\{\lambda_m\}_{m=1}^{\infty}$ of eigenvalues. Then in addition to the results in Theorem 4.1 the following results hold as well.
(a) The eigenvalues satisfy $\sum_{m=1}^{\infty} \lambda_m < \infty$.
(b) $\Gamma(\xi_1, \xi_2) = \sum_{m=1}^{\infty} \lambda_m \psi_m(\xi_1)\psi_m(\xi_2)$ for all $(\xi_1, \xi_2) \in \Xi \times \Xi$, where the right-hand side converges uniformly on $\Xi \times \Xi$, i.e.,

$$\lim_{n\to\infty} \sup_{(\xi_1,\xi_2)\in\Xi\times\Xi} \left| \Gamma(\xi_1, \xi_2) - \sum_{m=1}^{n} \lambda_m \psi_m(\xi_1)\psi_m(\xi_2) \right| = 0.$$

(c) The orthonormal sequence $\{\psi_m\}_{m=1}^{\infty}$ of eigenfunctions of Γ, including the eigenfunctions corresponding to zero eigenvalues, is complete in the Hilbert space $L^2(\mu)$, i.e., $L^2(\mu) = \mathrm{span}(\{\psi_m\}_{m=1}^{\infty})$.

Proof. Observe from the conditions of Theorem 5.1 that $\Gamma(\xi_1, \xi_2) \in L^2(\mu \times \mu)$. Moreover, with $\{\psi_m\}_{m=1}^{\infty}$ the sequence of the orthonormal eigenfunctions ψ_m of Γ, including the eigenfunctions corresponding to zero eigenvalues, if any, the space

$$S_2 = \mathrm{span}(\{\psi_k(\xi_1)\psi_m(\xi_2)\}_{k,m=1}^{\infty})$$

is a subspace of $L^2(\mu \times \mu)$. Then by the projection theorem (see Lemma 3.3) the projection of Γ on S_2 takes the form

$$\overline{\Gamma}(\xi_1, \xi_2) = \sum_{k=1}^{\infty}\sum_{k=1}^{\infty} c_{k,m}\psi_k(\xi_1)\psi_m(\xi_2)$$

where the $c_{k,m}$'s are such that $\int\int(\Gamma(\xi_1, \xi_2) - \overline{\Gamma}(\xi_1, \xi_2))^2\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2)$ is minimal. It is easy to verify, using (5.1), that this is the case for

$$\begin{aligned} c_{k,m} &= \int\int \psi_k(\xi_1)\Gamma(\xi_1, \xi_2)\psi_m(\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) \\ &= \lambda_m \int \psi_k(\xi_1)\psi_m(\xi_1)\mathrm{d}\mu(\xi_1) = \lambda_m I(k = m), \end{aligned}$$

where $I(.)$ is the indicator function. Hence, we can write

$$\Gamma(\xi_1,\xi_2) = \overline{\Gamma}(\xi_1,\xi_2) + K(\xi_1,\xi_2), \text{ where } \overline{\Gamma}(\xi_1,\xi_2) = \sum_{m=1}^{\infty} \lambda_m \psi_m(\xi_1)\psi_m(\xi_2) \quad (5.2)$$

and $K(\xi_1,\xi_2)$ is the projection residual, which by the projection theorem is orthogonal to S_2, i.e., for all $f \in S_2$, $\int\int K(\xi_1,\xi_2)f(\xi_1,\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = 0$. Thus, K is an element of the orthogonal complement $S_2^{\perp}$ of S_2, which in the present case is defined as

$$S_2^{\perp} = \left\{ g \in L^2(\mu\times\mu) : \int\int g(\xi_1,\xi_2)f(\xi_1,\xi_2)\mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) = 0 \text{ for all } f \in S_2 \right\}.$$

It is well-known (and easy to prove if not[12]) that $S_2^{\perp}$ is a Hilbert space itself, and if $S_2^{\perp} \neq \{0\}$ it has an orthonormal base of the type $\{\varphi_i(\xi_1)\varphi_j(\xi_2)\}_{i,j=1}^{M}$, $M \leq \infty$. See remark 5.1 below. For notational convenience, let us assume that $M = \infty$, so that

$$S_2^{\perp} = \operatorname{span}\left(\{\varphi_i(\xi_1)\varphi_j(\xi_2)\}_{i,j=1}^{\infty}\right)$$

and note that then

$$S_1^{\perp} = \operatorname{span}(\{\varphi_i\}_{i=1}^{\infty})$$

is the orthogonal complement of

$$S_1 = \operatorname{span}(\{\psi_m\}_{m=1}^{\infty}).$$

Moreover, note that

$$\begin{aligned} L^2(\mu) &= \operatorname{span}\left(\{\psi_m\}_{m=1}^{\infty}, \{\varphi_i\}_{i=1}^{\infty}\right), \\ L^2(\mu\times\mu) &= \operatorname{span}\left(\{\psi_k(\xi_1)\psi_m(\xi_2)\}_{k,m=1}^{\infty}, \{\varphi_i(\xi_1)\varphi_j(\xi_2)\}_{i,j=1}^{\infty}\right). \end{aligned}$$

Furthermore, note that $\int \overline{\Gamma}(\xi_1,\xi_2)\varphi_j(\xi_2)\mathrm{d}\mu(\xi_2) = 0$, hence

$$\int K(\xi_1,\xi_2)\varphi_j(\xi_2)\mathrm{d}\mu(\xi_2) = \int \Gamma(\xi_1,\xi_2)\varphi_j(\xi_2)\mathrm{d}\mu(\xi_2)$$

and thus by Lemma 3.1, treating $K(\xi_1,\xi_2)$ as a function of ξ_2 given ξ_1,

$$K(\xi_1,\xi_2) = \sum_{j=1}^{\infty} \langle \Gamma(\xi_1,.), \varphi_j(.)\rangle \varphi_j(\xi_2)$$

[12] By showing that every Cauchy sequence in $S_2^{\perp}$ takes a limit in $S_2^{\perp}$.

$$= \sum_{j=1}^{\infty} \left(\int \Gamma(\xi_1, \xi_*) \varphi_j(\xi_*) \mathrm{d}\mu(\xi_*) \right) \varphi_j(\xi_2) \tag{5.3}$$

$$= \sum_{j=1}^{\infty} \left(\int \Gamma(\xi_*, \xi_2) \varphi_j(\xi_*) \mathrm{d}\mu(\xi_*) \right) \varphi_j(\xi_1),$$

where the last equality follows by symmetry.

The series representation (5.3) is convenient in establishing that $K(\xi_1, \xi_2)$ is positive semidefinite, as follows. Let $f \in L^2(\mu)$ be arbitrary. Project f on S_1, with projection $f_1 \in S_1$ and residual $f_2 \in S_1^{\perp}$. By Lemma 3.1, f_1 and f_2 can be written as

$$f_1(\xi) = \sum_{m=1}^{\infty} \langle f_1, \psi_m \rangle \psi_m(\xi), \;\; f_2(\xi) = \sum_{j=1}^{\infty} \langle f_2, \varphi_j \rangle \varphi_j(\xi), \tag{5.4}$$

respectively. It is obvious from (5.3) and (5.4) that $\int K(\xi_1, \xi_2) f_1(\xi_2) \mathrm{d}\mu(\xi_2) = 0$, so that

$$\int\int f(\xi_1) K(\xi_1, \xi_2) f(\xi_2) \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2) = \int\int f_2(\xi_1) K(\xi_1, \xi_2) f_2(\xi_2) \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2).$$

It follows now from (5.3) and (5.4) that

$$\begin{aligned}
&\int\int f(\xi_1) K(\xi_1, \xi_2) f(\xi_2) \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2) \\
&= \int \left(\int f_2(\xi_1) \Gamma(\xi_1, \xi_*) \sum_{j=1}^{\infty} \varphi_j(\xi_*) \mathrm{d}\mu(\xi_*) \right) \left(\int \varphi_j(\xi_2) f_2(\xi_2) \mathrm{d}\mu(\xi_2) \right) \mathrm{d}\mu(\xi_1) \\
&= \int \left(\int f_2(\xi_1) \Gamma(\xi_1, \xi_*) \sum_{j=1}^{\infty} \langle f_2, \varphi_j \rangle \, \varphi_j(\xi_*) \mathrm{d}\mu(\xi_*) \right) \mathrm{d}\mu(\xi_1) \\
&= \int \left(\int f_2(\xi_1) \Gamma(\xi_1, \xi_*) f_2(\xi_*) \mathrm{d}\mu(\xi_*) \right) \mathrm{d}\mu(\xi_1) \\
&= \int\int f_2(\xi_1) \Gamma(\xi_1, \xi_2) f_2(\xi_2) \mathrm{d}\mu(\xi_1) \mathrm{d}\mu(\xi_2) \geq 0,
\end{aligned}$$

where the inequality follows from (3.3).

If $K(\xi_1, \xi_2)$ is also continuous on $\Xi \times \Xi$ then by the Hilbert-Schmidt theorem, K has at least one eigenfunction φ_0 with corresponding eigenvalue λ_0. However this is impossible because then φ_0 is also an eigenfunction of Γ and therefore is equal to one of the ψ_m's. Then the only possibility left is that $K(\xi_1, \xi_2) \equiv 0$ on $\Xi \times \Xi$.

To establish the continuity of $K(\xi_1, \xi_2)$, note that by the already established symmetric positive semidefiniteness of K,

$$K(\xi, \xi) \geq 0 \text{ on } \Xi,$$

and thus by (5.2),[13]

$$\sum_{m=1}^{\infty} \lambda_m \psi_m(\xi)^2 = \Gamma(\xi, \xi) - K(\xi, \xi) \leq \sup_{\xi \in \Xi} \Gamma(\xi, \xi) < \infty. \tag{5.5}$$

Integrating ξ out it follows that $\sum_{m=1}^{\infty} \lambda_m < \infty$, which proves part (a) of Theorem 5.1, and since $|\psi_m(\xi_1)\psi_m(\xi_2)| \leq 0.5(\psi_m(\xi_1)^2 + \psi_m(\xi_2)^2)$ it follows that

$$\sum_{m=1}^{\infty} \lambda_m |\psi_m(\xi_1)\psi_m(\xi_2)| \leq \sup_{\xi \in \Xi} \Gamma(\xi, \xi) < \infty. \tag{5.6}$$

The latter implies that

$$\begin{aligned} \infty > \alpha_{n+1} &= \sup_{(\xi_1, \xi_2) \in \Xi \times \Xi} \sum_{m=n+1}^{\infty} \lambda_m |\psi_m(\xi_1)\psi_m(\xi_2)| \\ &\geq \lambda_{n+1} |\psi_{n+1}(\xi_1^*)\psi_{n+1}(\xi_2^*)| + \alpha_{n+2} \end{aligned}$$

for all $(\xi_1^*, \xi_2^*) \in \Xi \times \Xi$, hence

$$\lambda_{n+1} \sup_{(\xi_1, \xi_2) \in \Xi \times \Xi} |\psi_{n+1}(\xi_1)\psi_{n+1}(\xi_2)| \leq \alpha_{n+1} - \alpha_{n+2}$$

Moreover, since α_n is non-increasing, finite and non-negative, $\lim_{n \to \infty} \alpha_n = \alpha$ exists. Furthermore, for $N > n$,

$$\sum_{m=n+1}^{N} \lambda_m \sup_{(\xi_1, \xi_2) \in \Xi \times \Xi} |\psi_m(\xi_1)\psi_m(\xi_2)| \leq \sum_{m=n+1}^{N} (\alpha_m - \alpha_{m+1}) = \alpha_{n+1} - \alpha_{N+1}.$$

Hence, letting $N \to \infty$ first and then $n \to \infty$, it follows that

$$\lim_{n \to \infty} \sum_{m=n+1}^{\infty} \lambda_m \sup_{(\xi_1, \xi_2) \in \Xi \times \Xi} |\psi_m(\xi_1)\psi_m(\xi_2)| \leq \lim_{n \to \infty} \alpha_{n+1} - \alpha = 0.$$

[13] This paragraph is based on a related idea of Mercer (1909).

Consequently,

$$\lim_{n\to\infty}\sup_{(\xi_1,\xi_2)\in\Xi\times\Xi}\left|\overline{\Gamma}(\xi_1,\xi_2)-\sum_{m=1}^{n}\lambda_m\psi_m(\xi_1)\psi_m(\xi_2)\right|$$
$$\leq \lim_{n\to\infty}\sum_{m=n+1}^{\infty}\lambda_m\sup_{(\xi_1,\xi_2)\in\Xi\times\Xi}|\psi_m(\xi_1)\psi_m(\xi_2)|=0. \tag{5.7}$$

It follows now easily from the continuity of $\sum_{m=1}^{n}\lambda_m\psi_m(\xi_1)\psi_m(\xi_2)$ that $\overline{\Gamma}(\xi_1,\xi_2)$ is continuous, hence $K(\xi_1,\xi_2)=\Gamma(\xi_1,\xi_2)-\overline{\Gamma}(\xi_1,\xi_2)$ is continuous.

However, as argued before, continuity of K implies that $K(\xi_1,\xi_2)\equiv 0$ on $\Xi\times\Xi$, and thus $\Gamma(\xi_1,\xi_2)\equiv\overline{\Gamma}(\xi_1,\xi_2)$ on $\Xi\times\Xi$. The latter together with (5.7) imply that part (b) of Theorem 5.1 holds.

Now any element $\varphi\in S_1^{\perp}=\text{span}(\{\varphi_i\}_{i=1}^{\infty})$ is an eigenfunction of Γ, hence $\varphi\in S_1=\text{span}(\{\psi_m\}_{m=1}^{\infty})$ and thus $S_1^{\perp}\subset S_1$. The latter is only possible if $S_1^{\perp}=\{0\}$. Consequently, the orthonormal sequence $\{\varphi_i\}_{i=1}^{\infty}$ does not exist, and therefore $L^2(\mu)=\text{span}(\{\psi_m\}_{m=1}^{\infty}\}$. This proves part (c) of Theorem 5.1. ■

Remark 5.1. Suppose that the sequence $\{\psi_m\}_{m=1}^{\infty}$ is not complete in $L^2(\mu)$. Then the orthogonal complement $S_1^{\perp}$ of $S_1=\text{span}(\{\psi_m\}_{m=1}^{\infty})$ is larger than $\{0\}$. To construct a complete orthonormal sequence $\{\varphi_j\}_{j=1}^{M}$, $M\leq\infty$, in $S_1^{\perp}$, recall from Lemma 3.1 that there exists an orthonormal basis $\{\eta_m\}_{m=1}^{\infty}$ of $L^2(\mu)$, i.e., $L^2(\mu)=\text{span}(\{\eta_m\}_{m=1}^{\infty})$. Now for each m, project η_m on S_1, and let r_m be the residual involved. Note that $r_m\in S_1^{\perp}$. Let m_j be the subsequence for which $||r_{m_j}||>0$, and $r_m=0$ for all other m's. Denote $\overline{r}_j=r_{m_j}$. Then $S_1^{\perp}=\text{span}(\{\overline{r}_j\}_{j=1}^{\infty})$. However, the sequence $\{\overline{r}_j\}_{j=1}^{\infty}$ may not be orthonormal. To make it orthogonal, let $\varphi_1=||\overline{r}_1||^{-1}\overline{r}_1$, and construct the orthonormal sequence $\{\varphi_j\}_{j=1}^{\infty}$ recursively as follows. For $j\geq 2$, project $\overline{r}_j$ on $\text{span}(\{\varphi_m\}_{m=1}^{j-1})$. The residual of this projection takes the form $\overline{r}_j^*=\overline{r}_j-\sum_{m=1}^{j-1}\langle\overline{r}_j,\varphi_m\rangle\varphi_{m.}$ Setting $\varphi_j=||\overline{r}_j^*||^{-1}\overline{r}_j^*$ and repeating this construction for $j=2,3,...,M$ then yields an orthonormal sequence $\{\varphi_j\}_{j=1}^{M}$ such that $S_1^{\perp}=\text{span}(\{\varphi_j\}_{j=1}^{M})$. The latter implies that $S_2^{\perp}=\text{span}(\{\varphi_i(\xi_1)\varphi_j(\xi_2)\}_{i,j=1}^{M})$.

Remark 5.2. There have been various attempts in the literature to generalize Mercer's theorem to unbounded domains Ξ. See for example Steinwart and Scovel (2012) and the references therein. In particular, Steinwart and Scovel state that "all extensions of Mercers theorem in this direction either stick too closely to the original topological structure of Ξ and Γ, or replace the absolute and uniform convergence by weaker notions of convergence that are not strong enough for many statistical applications." On the other hand, the compactness of Ξ is used in the proof of Theorem 4.1 only to guarantee that

$$\int\int \Gamma\left(\xi_1,\xi_2\right)^2 \mathrm{d}\mu(\xi_1)\mathrm{d}\mu(\xi_2) < \infty, \tag{5.8}$$

and in the proof of Theorem 5.1 the compactness of Ξ is only used in (5.5) and (5.6), together with the continuity of $\Gamma(\xi_1,\xi_2)$ on $\Xi\times\Xi$, to guarantee that $\sup_{\xi\in\Xi}\Gamma(\xi,\xi) < \infty$. Therefore, Mercer's theorem carries over to probability measures μ on unbounded domains Ξ as long as (5.8) holds and $\Gamma(\xi,\xi)$ is uniformly bounded.

6. Bootstrap

Recall that for real valued weight functions $w_t(\xi)$ the ICM test statistic takes the form $\widehat{T}_n = \int \widehat{Z}_n(\xi)^2\mathrm{d}\mu(\xi)$, where

$$\widehat{Z}_n(\xi) = n^{-1/2}\sum_{t=1}^{n}\widehat{U}_t w_t(\xi).$$

Moreover, denoting

$$Z_n(\xi) = n^{-1/2}\sum_{t=1}^{n} U_t\phi_t(\xi)$$

where

$$\begin{aligned}
\phi_t(\xi) &= w_t(\xi) - b(\xi)'A^{-1}\nabla f_t(\theta_0),\\
b(\xi) &= E\left[w_t(\xi)\nabla f_t(\theta_0)\right],\ A = E\left[(\nabla f_t(\theta_0))(\nabla f_t(\theta_0))'\right],\\
\nabla f_t(\theta) &= (\partial/\partial\theta')f_t(\theta),
\end{aligned}$$

it follows that under the null hypothesis and the conditions in BP,

$$\widehat{Z}_n(\xi) = Z_n(\xi) + o_p(1) \Rightarrow Z_0(\xi),$$

hence

$$\widehat{T}_n \xrightarrow{\mathrm{d}} T_0 = \int Z_0(\xi)^2\mathrm{d}\mu(\xi)$$

where $Z_0(\xi)$ is a zero mean Gaussian process on Ξ with covariance function

$$\Gamma(\xi_1,\xi_2) = E[U_t^2\phi_t(\xi_1)\phi_t(\xi_2)].$$

Next, let for $b = 1, 2, ..., M$,

$$\widetilde{Z}_{b,n}(\xi) = n^{-1/2} \sum_{t=1}^{n} \varepsilon_{b,t} \widehat{U}_t \widehat{\phi}_{n,t}(\xi),$$

where the $\varepsilon_{b,t}$'s are independent random drawings from the standard normal distribution, and

$$\begin{aligned} \widehat{\phi}_{n,t}(\xi) &= w_t(\xi) - \widehat{b}_n(\xi)' \widehat{A}_n^{-1} \nabla f_t(\widehat{\theta}), \text{ with} \\ \widehat{b}_n(\xi) &= \frac{1}{n} \sum_{t=1}^{n} w_t(\xi) \nabla f_t(\widehat{\theta}), \ \widehat{A}_n = \frac{1}{n} \sum_{t=1}^{n} (\nabla f_t(\widehat{\theta}))(\nabla f_t(\widehat{\theta}))'. \end{aligned}$$

The $\widetilde{Z}_{b,n}(\xi)$'s are the (wild) bootstrap versions of $Z_n(\xi)$. Then *conditional on the data*, the $\widetilde{Z}_{b,n}(\xi)$'s are zero-mean Gaussian processes with covariance function

$$\widetilde{\Gamma}_n(\xi_1, \xi_2) = \frac{1}{n} \sum_{t=1}^{n} \widehat{U}_t^2 \widehat{\phi}_{n,t}(\xi_1) \widehat{\phi}_{n,t}(\xi_2).$$

It is easy to verify that under the conditions in BP, $\widetilde{Z}_{b,n} \Rightarrow Z_b$, where Z_b has the same distribution as Z_0. Moreover, since conditional on the data-generating process the $Z_b(\xi)$'s are independent (because the $\varepsilon_{b,t}$'s are independent of the data), it follows that the Z_b's are unconditionally independent, and independent of Z_0 as well. Therefore, the following result holds.

Theorem 6.1. *For $b = 1, 2, ..., M$ with M fixed, denote $\widetilde{T}_{b,n} = \int \widetilde{Z}_{b,n}(\xi)^2 \mathrm{d}\mu(\xi)$, and recall that $\widehat{T}_n = \int \widehat{Z}_n(\xi)^2 \mathrm{d}\mu(\xi)$. Under H_0, the conditions in BP and the additional condition $E[\sup_{\theta \in \Theta} (Y_t - f_t(\theta))^2 ((\partial/\partial\theta_{i_1})(\partial/\partial\theta_{i_2}) f_t(\theta))^2] < \infty$ for $i_1, i_2 = 1, 2, ..., m$,*

$$\left(\widehat{T}_n, \widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}\right) \xrightarrow{\mathrm{d}} (T_0, T_1, T_2, ..., T_M)$$

where for $i = 0, 1, ..., M$ the T_i's are i.i.d. Moreover, $\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}$ have the same distributions and conditional on the data they are independent.

Proof. Similar to Theorem 6.2 in the addendum to Bierens (1982) in Chapter 2. ■

Sorting $\widetilde{T}_{1,n}, \widetilde{T}_{2,n}, ..., \widetilde{T}_{M,n}$ in *decreasing* order, the $\alpha \times 100\%$ bootstrap critical value is[14]

$$\widetilde{c}_{M,n}(\alpha) = \widetilde{T}_{[\alpha M],n},$$

which approximates the asymptotic critical value $c_0(\alpha)$, i.e.,

$$\Pr[T_0 > c_0(\alpha)] = \alpha,$$

in the following way. For an arbitrary $\varepsilon > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \Pr[|\widetilde{c}_{M,n}(\alpha) - c_0(\alpha)| > \varepsilon] = 0.$$

See section 6 in the addendum to Bierens (1982) in Chapter 2.

Note that this bootstrap procedure is an adaptation to the ICM test of the bootstrap procedure for threshold regressions proposed by Hansen (1996).

7. Concluding remarks

In the mathematical literature the Hilbert-Schmidt theorem and Mercer' theorem are nowadays usually derived as by-products of linear operator theory,[15] which however is way over my head. Therefore, in this addendum to BP I have presented alternative proofs which only require elementary Hilbert space theory. I do not claim any originality. May be these proofs are already done in this way, but if so I am not aware of any references. The only originality claim I can make is that I figured out these proofs all by myself.

In BP, Mercer's theorem (Lemma 1) was used to show that the ICM test has nontrivial $\sqrt{n}$ local power and is admissible. Moreover, Lemma 1 was also used to derive upper bounds of the critical values. However, Mercer's theorem does not play a role for the weak convergence result of the empirical process $\widehat{Z}_n(\xi) = n^{-1/2}\sum_{t=1}^{n}\widehat{U}_t w_t(\xi)$, and neither does the bootstrap procedure proposed in this addendum. Therefore, as far as the global power and size of the ICM test are concerned, it seems possible to generalize the ICM test further by allowing the probability measure μ to be the induced probability measure of an absolutely continuous distribution on $\mathbb{R}^k$ instead of the uniform probability measure on a compact set $\Xi \subset \mathbb{R}^k$. If so, we no longer need to transform the conditioning

[14] Here $[x]$ denotes the largest integer $\leq x$.

[15] See for example Young (1988, Ch. 8), Dunford and Schwartz (1963), Krein (1998) and Steinwart and Scovel (2012), among others.

variables in the weight function $\phi_t(\xi)$ to bounded variables. On the other hand, Boning and Sowell (1999) have shown that, given the compact set Ξ, the uniform probability measure is optimal in the sense that then the ICM test has the greatest weighted average local power as defined in Andrews and Ploberger (1994).

Finally, it should be noted that the ICM test in BP and its generalization in this addendum is only consistent in the i.i.d. case. In the time series case the correctness of the regression model specification involved is equivalent to the hypothesis that the errors U_t are martingale differences relative to an appropriate filtration $\mathcal{F}_t$. Since in BP the dimension of the vector X_t of conditioning variables is fixed, it is possible that $E[U_t|X_t] = 0$ a.s. while $\Pr(E[U_t|\mathcal{F}_{t-1}] = 0) < 1$. De Jong (1996) remedies this problem by allowing the dimension of X_t to increase to infinity with the sample size, whereas earlier in Bierens (1984) I have proposed to use a weighted average of ICM test statistics $\widehat{T}_k$ based on conditioning vectors $X_{k,t} \in \mathbb{R}^k$. Applying the results in BP and in this addendum in combination with the weighted ICM testing approach in Bierens (1984) then yields a genuine consistent test of time series regression models. See the addendum to Bierens (1984) in Chapter 3.

References

Andrews, D. W. K., and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter is Present only Under the Alternative", *Econometrica* 62, 1383-1414.

Bernkopf, M. (1966): "The Development of Function Spaces with Particular Reference to their Origins in Integral Equation Theory", *Archive for History of Exact Sciences* 3, 1-96.

Bierens, H. J. (1982): "Consistent Model Specification Tests", *Journal of Econometrics* 20, 105-134, reprinted in Chapter 2.

Bierens, H. J. (1984): "Model Specification Testing of Time Series Regressions", *Journal of Econometrics* 26, 323-353, reprinted in Chapter 3.

Bierens, H. J. (1990): "A Consistent Conditional Moment Test of Functional Form", *Econometrica* 58, 1443-1458, reprinted in Chapter 4.

Bierens, H. J. (1994): "Comment on 'Artificial Neural Networks: An Econometric Perspective' by Chung-Ming Kuan and Halbert White", *Econometric Reviews* 13, 93-97.

Bierens, H. J. (2004): *Introduction to the Mathematical and Statistical Foundations of Econometrics*, Cambridge University Press.

Bierens, H. J. (2014): "The Hilbert Space Theoretical Foundation of Semi-Nonparametric Modeling", Chapter 1 in: J. Racine, L. Su and A. Ullah (eds), *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press.

Bierens, H. J., and W. Ploberger (1997): "Asymptotic Theory of Integrated Conditional Moment Tests", *Econometrica* 65, 1129-1151, reprinted in Chapter 5.

Boning, W. B., and F. Sowell (1999): "Optimality for the Integrated Conditional Moment Test", *Econometric Theory* 15, 710-718.

Chung, K. L. (1974): *A Course in Probability Theory*, Academic Press.

de Jong, R. (1996): "The Bierens Test Under Data Dependence", *Journal of Econometrics* 72, 1-32.

Dunford, N., and J. T . Schwartz (1963): *Linear Operators, Part II: Spectral Theory*, Wiley Interscience.

Hansen, B. E. (1996): "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis", *Econometrica* 64, 413-430.

Krein, M. G. (1998): "Compact Linear Operators on Functional Spaces with Two Norms", *Integral Equations and Operator Theory* 30, 140-162.

Kuan, C. M. (2008): "Artificial Neural Networks", in S. N. Durlauf and L. E. Blume (eds.), *New Palgrave Dictionary of Economics*, Palgrave Macmillan.

Kuan, C. M., and H. White (1994): "Artificial Neural Networks: An Econometric Perspective", *Econometric Reviews* 13, 1-92.

Mercer, J. (1909): "Functions of Positive and Negative Type and their Connection with the Theory of Integral Equations", *Philosophical Transactions of the Royal Society A* 209, 415-446.

Siegmund-Schultze, R. (1986), "Der Beweis des Hilbert-Schmidt-Theorems", *Archive for History of Exact Sciences* 36, 251-270.

Steinwart, I. and C. Scovel (2012), "Mercers Theorem on General Domains: On the Interaction between Measures, Kernels, and RKHSs", *Constructive Approximation* 35, 363-417.

Stinchcombe, M. B., and H. White (1998): "Consistent Specification Testing with Nuisance Parameters Present Only Under the Alternative", *Econometric Theory* 14, 295-325.

Young, N. (1988): *An Introduction to Hilbert Space*, Cambridge University Press.