J. Appl. Prob. Spec. Vol. **51A,** 189–201 (2014)

CELEBRATING 50 YEARS OF THE APPLIED PROBABILITY TRUST

Edited by
S. ASMUSSEN, P. JAGERS, I. MOLCHANOV and L. C. G. ROGERS

Part 5. Finance and econometrics

ASPECTS OF PREDICTION

N. H. BINGHAM, *Imperial College London*
Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Email address: n.bingham@imperial.ac.uk

BADR MISSAOUI, *Imperial College London*
Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK. Email address: badr.missaoui08@imperial.ac.uk

APPLIED PROBABILITY TRUST
DECEMBER 2014

ASPECTS OF PREDICTION

BY N. H. BINGHAM AND BADR MISSAOUI

Abstract

We survey some aspects of the classical prediction theory for stationary processes, in discrete time in Section 1, turning in Section 2 to continuous time, with particular reference to reproducing-kernel Hilbert spaces and the sampling theorem. We discuss the discrete-continuous theories of ARMA-CARMA, GARCH-COGARCH, and OPUC-COPUC in Section 3. We compare the various models treated in Section 4 by how well they model volatility, in particular volatility clustering. We discuss the infinite-dimensional case in Section 5, and turn briefly to applications in Section 6.

Keywords: Stationary process; Kolmogorov isomorphism theorem; time series; stochastic volatility; volatility clustering; Banach space; covariance operator; locally convex; reproducing-kernel Hilbert space; sampling theorem

2010 Mathematics Subject Classification: Primary 60-02
Secondary 62-02

1. Prediction theory of stationary sequences

We begin with the most classical aspect: the prediction theory of stationary stochastic processes $X = (X_n)$ in discrete time n and one dimension (we take X to be complex valued). There is a good account of this subject up to 1958 in [39], and a great deal of relevant recent work in the books of Simon [91, 92, 93]. We use as our main reference here the recent surveys by the first author [10, 11] on the probabilistic side, together with its statistical sequel [12].

We take X to have zero mean and finite variance, and to be (wide-sense) stationary, with autocovariance function $\gamma = (\gamma_n)$, $\gamma_n = \mathbb{E}[X_n\overline{X}_0]$ (the variance is constant by stationarity, so we may take it as 1 for convenience). Let $\mathcal{H}$ be the Hilbert space spanned by $X = (X_n)$ in the L_2-space of the underlying probability space, with inner product $(X, Y) := \mathbb{E}[X\overline{Y}]$ and norm $\|X\| := [\mathbb{E}(|X|^2)]^{1/2}$. In the complex plane, we write $\mathbb{T}$ for the unit circle, the boundary of the unit disc $\mathbb{D}$, parameterised by $z = e^{i\theta}$; unspecified integrals are over $\mathbb{T}$.

1.1. The Cramér representation and Kolmogorov isomorphism theorem

There are four key ingredients here.

(i) There is a process Y on $\mathbb{T}$ with orthogonal increments with Fourier coefficients X_n; by the *Cramér representation,*

$$X_n = \int e^{in\theta}\, Y(d\theta) \quad \text{for all } n \in \mathbb{Z} \tag{CR}$$

(see [23], [24, Sections 5.6, 5.7, 7.5]; cf. [28, Section X.4]); Y is called the *Cramér process*. One can regard Y as a random measure, in which case Y is *orthogonally scattered* (the term is due to Masani [72]).

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

(ii) There is a measure μ on $\mathbb{T}$, the *spectral measure*, with

$$\mathbb{E}[Y(A)\overline{Y(B)}] = \mu(A \cap B).$$

In differential notation, this can be written as

$$\mathbb{E}[\mathrm{d}Y(\theta)^2] = \mathrm{d}\mu(\theta), \tag{SM}$$

which shows the link with Itô calculus. See, e.g. [28, Sections IX.1, X.4].

(iii) The autocorrelation function γ then has the *spectral representation*

$$\gamma_n = \int \mathrm{e}^{-in\theta}\, \mu(\mathrm{d}\theta) \quad \text{for all } n \in \mathbb{Z} \tag{SR}$$

(Herglotz's theorem of 1911).

(iv) We have the *Kolmogorov isomorphism* between $\mathcal{H}$ (the *time domain*) and $L_2(\mu)$ (the *frequency domain*) given by

$$X_t \leftrightarrow \mathrm{e}^{\mathrm{i}t\cdot} : \theta \in \mathbb{R}/2\pi\mathbb{Z} \mapsto \mathrm{e}^{\mathrm{i}\theta\cdot} \in \mathbb{T} \tag{KIT}$$

(see [59]) for integer t (as time is discrete). A good historical account of Kolmogorov's work in this area is given in [90, 'The forties (1940–49)']. For modern textbook accounts of the background, see, e.g. [49, Chapter 13 including Notes] and [53, Chapter 7.4]. Stochastic processes expressible as the Fourier transform of another process are called *harmonizable* (the term is due to Loève [65], [66, Section 37.4]); see [54, 83].

To exclude trivialities, it is necessary to take μ to have infinite support (to be *nontrivial*: [91, p. 1]); we shall do so in what follows, without further comment.

For reasons of space, we omit state space models and the Kalman filter; see, e.g. [19, Chapter 12].

1.2. Verblunsky's theorem and the partial autocorrelation function

By the Kolmogorov isomorphism theorem, all the information relevant to the time evolution of the process is encoded in the spectral measure μ (though other relevant information, such as the distribution of the X_n, is not; see below). Now to model μ, we have to proceed nonparametrically, as the situation is infinite dimensional, and it might be preferable to use a sequence representation rather than a measure (or density-function) representation. This option is available to us, thanks to *Verblunsky's theorem* (S. Verblunsky in 1935 and 1936; the result is named and proved in [91], to which we refer for details), according to which we have a bijection (the *Verblunsky bijection*) between the spectral measures μ (probability measures on $\mathbb{T}$) and sequences $\alpha = (\alpha_n)_{n\in\mathbb{Z}}$, with each $\alpha_n \in \mathbb{D}$ (α is the *Verblunsky sequence*, or *partial autocorrelation function* (PACF), and the α_n are the *Verblunsky coefficients*). Verblunsky discovered this result in analysis; it was rediscovered much later in statistics by Barndorff-Nielsen and Schou in 1973 and Ramsey in 1974. It is extremely useful in statistics as it provides an *unrestricted parameterisation* (all sequences with terms in the unit disc are possible; this is not the case for the autocorrelation function (ACF) γ). Then α_n is the correlation between the residuals of X_0 and X_n when regressed on the values $X_1, \ldots, X_{n-1}$ in between. Using the usual notation for Hilbert-space projections,

$$\alpha_n := \mathrm{corr}(X_n - P_{[1,n-1]}X_n,\, X_0 - P_{[1,n-1]}X_0);$$

see, e.g. [10] for details and references.

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

1.3. Orthogonal polynomials on the unit circle

In Szegö's classic book [94] on orthogonal polynomials (on the real line: OPRL), one chapter, Chapter XI, is devoted to orthogonal polynomials on the unit circle (OPUC). Although Verblunsky is not mentioned in [94], his work is the guiding theme in the two-volume book [91, 92]; we summarise here what we need. Recall from OPRL the key role played by the three-term recurrence relation (Favard's theorem); this involves two sequences of coefficients. In OPUC, there is again a three-term recurrence relation, the *Szegö recursion*, but now only one sequence of coefficients (the Verblunsky coefficients above). For each probability measure μ on $\mathbb{T}$ (the spectral measure above), write P_n for the monic orthogonal polynomials they generate (by Gram–Schmidt orthogonalization). For every polynomial Q_n of degree n, write

$$Q_n^*(z) := z^n \overline{Q_n\left(\frac{1}{\bar{z}}\right)}$$

for the *reversed polynomial*. Then the Szegö recursion is

$$P_{n+1}(z) = zP_n(z) - \bar{\alpha}_{n+1}P_n^*(z).$$

Here the parameters α_n are the Verblunsky coefficients, and lie in $\mathbb{D}$ by Verblunsky's theorem ($|\alpha_n| \leq 1$ as correlations lie in $\mathbb{D}$ automatically; that $|\alpha_n| < 1$ follows from μ being nontrivial).

The Szegö recursion is known in the time-series literature as the *Levinson–Durbin algorithm*. The best linear predictor of the immediate future based on the present and a finite segment of the past is a linear combination of the given values; their coefficients form a triangular matrix $A = (a_{nk})$; the Verblunsky coefficients or PACF are the diagonal elements in A. See, e.g. [19, Sections 3.4, 5.2] and [10, Section 3].

1.4. The Wold decomposition

Write w for the *spectral density*. Then $w/2\pi$ is the density of the absolutely continuous component of μ:

$$\mu = \mu_s + \mu_{ac} = \mu_s + \frac{w\,d\theta}{2\pi}.$$

Write σ^2 for the one-step mean-square prediction error:

$$\sigma^2 := \mathbb{E}[(X_0 - P_{(-\infty,-1]}X_0)^2].$$

Call X *nondeterministic* if $\sigma > 0$, and *deterministic* if $\sigma = 0$. The *Wold decomposition* of 1938 [24, Section 5.7], [19, Section 5.7], [10, Section 3] expresses a process X as the sum of a non-deterministic process U (the 'good' part) and a deterministic process V (the 'bad' part), namely,

$$X_n = U_n + V_n,$$

where the process U is a moving average,

$$U_n = \sum_{j=-\infty}^{n} m_{n-j}\xi_j = \sum_{k=0}^{\infty} m_k \xi_{n-k},$$

in which the ξ_j, the *innovations* at time n, have zero mean and are uncorrelated both with each other and with $V = (V_n)$, i.e. $\mathbb{E}[\xi_n] = 0$ and $\mathrm{var}(\xi_n) = \mathbb{E}[\xi_n^2] = \sigma^2$. Thus, when $\sigma = 0$, the ξ_n are 0, U is missing, and the process is deterministic. When $\sigma > 0$, the spectral measures of U_n and V_n are μ_{ac} and μ_s, the absolutely continuous and singular components of μ (again, the 'good' and 'bad' parts). Then the Wold decomposition agrees with the Lebesgue decomposition

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

of the spectral measure; in view of the Cramér representation, this is often called the *Wold–Cramér concordance*. It fails in dimension higher than 1; see Section 5.

1.5. Szegö's theorem

For a process genuinely evolving in time, one expects new information as time passes, i.e. when $\sigma > 0$, there is nondeterminism. The condition for this is *Szegö's condition*; see (Sz) below. Szegö's theorem gives the equivalence of the following:

(i) $\sigma > 0$;

(ii) $\log w \in L_1(\mathbb{T})$, i.e. (as $\int \log w < \infty$ by Jensen's inequality)

$$\int \log w(\theta)\,\mathrm{d}\theta > -\infty; \tag{Sz}$$

(iii) $\alpha \in \ell_2$.

Then $\sigma^2 = \prod_1^\infty (1 - |\alpha_n|^2)$, and from Kolmogorov's formula [59], σ^2 is the geometric mean $G(\mu)$ of μ:

$$\sigma^2 = \exp\left(\frac{1}{2\pi}\int \log w(\theta)\,\mathrm{d}\theta\right) =: G(\mu) > 0 \tag{K}$$

(see, e.g. [10, Section 4] for details and references). (The Szegö condition is visibly of entropy type, and is related to the Gibbs variational principle of statistical mechanics.)

1.6. The Szegö function

We now restrict attention to when (Sz) holds. Then the *Szegö function*

$$h(z) := \exp\left(\frac{1}{4\pi}\int \frac{\mathrm{e}^{\mathrm{i}\theta}+z}{\mathrm{e}^{\mathrm{i}\theta}-z}\log w(\theta)\,\mathrm{d}\theta\right), \qquad z \in D, \tag{OF}$$

exists, and is an *outer function* (hence (OF)) in the Hardy space $H_2(\mathbb{T})$ on the torus $\mathbb{T}$ (see, e.g. [10, Section 4]). It is an 'analytic square root' of w:

$$|h(\mathrm{e}^{\mathrm{i}\theta})|^2 = w(\theta), \qquad h = [w^{1/2}]$$

(see [76] and [77, p. 380]). We have (see [46] and [47, Section 2] for background) $h(z) = \sum_{n=0}^\infty m_n z^n$: the Maclaurin coefficients $m = (m_n)$ of h are the moving average (MA(∞)) coefficients in the Wold decomposition. Those of $-1/h(z) = \sum_{n=0}^\infty r_n z^n$, $z \in D$, give the AR(∞) coefficients $r = (r_n)$ in the (infinite-order) *autoregression* $\sum_{j=-\infty}^n r_{n-j} X_j + \xi_n = 0$, $n \in Z$.

2. Continuous time

As always, one has to choose from context between modelling our process in discrete or continuous time. Such matters are topical: the ARMA and GARCH models just mentioned have led on to extensive work on continuous-time versions such as CARMA and COGARCH.

Within the context of econometrics: data are discrete, and as a result, the discrete-time theory is the more developed. But continuous time offers advantages, such as a setting in which one can use calculus, and model the dynamics directly. The classic text in this area is that of Bergstrom [9]. Continuous-time econometrics, and finance, have flourished as a result of the extensive use of stochastic calculus, in Black–Scholes(–Merton) theory, etc., as the title of [50] bears witness.

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

First, we have the obvious but strikingly useful fact that the Cramér representation (CR) of Section 1 allows us to pass directly to continuous time, simply by replacing the discrete time n there by a continuous time t:

$$X_t = \int e^{it\theta}\, Y(d\theta), \qquad t \in \mathbb{R}. \tag{CR}$$

This emphasises dramatically the key structural role played here by the Cramér process Y. Of course, this procedure commits us to a choice of unit of time; it works best where the context has a natural one.

Note first the nature of the process $X = (X_t)$ thus defined. It is *harmonizable* (Section 1); see [64, 75] for early work on such harmonizable processes. In engineering language, the compact support $[-\pi, \pi]$ of Y makes Y *band limited*, whence X_t is a random analytic function of t (see [8] and [24, Section 7.3]). By the Paley–Wiener theorem (see [56, Chapter VI.7] and [78]), X is a *random entire function*, of *exponential type* π.

2.1. The sampling theorem, Paley–Wiener spaces, and reproducing-kernel Hilbert spaces

The Paley–Wiener spaces PW_a (with general $a > 0$ rather than $a = \pi$ as here) are the motivating examples of *de Branges spaces* [26], [31, Section 6.1]. They are also classical examples of *reproducing-kernel Hilbert spaces* (RKHS) [77, B, Section 6.5.2]. What is needed for a Hilbert space to have a reproducing kernel is that the point-evaluation maps be continuous; thus, Hilbert spaces of continuous (and, in particular, analytic) functions have reproducing kernels. The reproducing kernel for PW_a is the sinc *function* (sinus cardinalis: this stems from E. T. Whittaker's work of 1915; the name cardinal series is from J. M. Whittaker in the 1920s; the name sinc is from P. M. Woodward in 1953). For $f \in PW_\pi$, we have

$$f(x) = \sum_{n\in\mathbb{Z}} f(n)\,\mathrm{sinc}(x-n) \quad \text{for all } x \in \mathbb{R}, \qquad \mathrm{sinc}(x) := \frac{\sin \pi x}{\pi x}.$$

This recovery of a function everywhere on the line from its values sampled at the integers is called the *sampling theorem* (or Whittaker–Shannon–Kotelnikov sampling theorem: J. M. Whittaker [97] in 1935, C. E. Shannon in 1949, and V. I. Kotelnikov in 1933). For background, see, e.g. [42, Section 2.6.3], [43, 44, 45], and [79, Section 7.2].

This works more generally. That any RKHS H with kernel K with a total orthogonal set of point evaluation vectors δ_{x_n} has the sampling property

$$f(x) = \sum_{n\in\mathbb{Z}} f(x_n) K(x, x_n), \qquad f \in H, \tag{SP}$$

is Kramer's sampling theorem [61]. Martin [71] considered such results from the point of view of operator theory, and showed in particular that many de Branges spaces have this property.

The sampling theorem is clearly closely connected with the Paley–Wiener theorem from the above, and is in fact an easy corollary of it (see [31, Section 2.2]); similarly for the Poisson summation formula (see [14]). Butzer *et al.* [21] showed the equivalence of all three, and of the general Parseval formula, in the band-limited case.

2.2. The Kolmogorov–Wiener filter

The prediction problem given the infinite past was studied in continuous time by Kolmogorov [59] in 1941 (see [90]), M. G. Krein in 1944, and also by Wiener [98]; see [28, Chapter XII]. Under (Sz), as before, the Szegö function h exists, and is in the Hardy space H_2 on the upper

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

half-plane; again, $w = |h|^2$. With $\hat{h}$ the Fourier transform of h, the mean-square error of the best linear predictor of X_T given $\{X_t : t \le 0\}$ is $\int_0^T |\hat{h}(t)|^2 \, dt$ [30, Section 5].

Wiener's work was done, independently of Kolmogorov, during 1940–41, restricted for security reasons during World War II, circulated subject to this (as the 'Yellow Book'), and published in 1949. It contains (pages 55, 59) Wiener's detailed account of the connections between his work and Kolmogorov's, and also two mathematical appendices by Wiener's colleague and former pupil Levinson. Volume III of Wiener's *Collected Works* [99] contains thirteen of his papers on prediction (published 1949–59; six with Masani, one with Akutowicz), together with commentaries (by Kailath, Akutowicz, Masani, Salehi, Muhly, and Kallianpur). These papers contain further discussion by Wiener of the connections between his work and Kolmogorov's, and between discrete and continuous time.

2.3. Prediction given a finite past

The continuous-time prediction problem given a finite segment of the past is harder, and has been studied by Krein (1944–54), and Dym and McKean [29, 30, 31]. Both methods use the Kolmogorov isomorphism $X_t \leftrightarrow e^{it\cdot}$ of Section 1 to map the given segment $\{X_t : -T \le t \le T\}$ to the Paley–Wiener space PW_T. Tools used include Krein's theory of strings and results of Levinson and McKean [62] (this work has also been found useful in discrete time [12, Section 7.4], [55]). Matters are too technical to attempt a summary here; for details, see [29], [30, Sections 7–23], and [31, Chapter 4].

3. ARMA/CARMA, GARCH/COGARCH, OPUC/COPUC

The general stationary process is an infinite-dimensional object (parameterised several different ways in Section 1), but in practice one has to truncate and work finite dimensionally. One familiar way to do this is to use ARMA (autoregressive moving average) models; see, e.g. [16] and [19, Chapters 3, 4, 8]. This is flexible and practicable, but the $p+q$ parameters of the ARMA(p, q) model chosen may have little or no meaning in reality.

As in Section 2, one may prefer to work in continuous time (CARMA). Here the polynomials in the ARMA(p, q) model in the lag operator (equivalently, the (forward or backward) difference operator) are replaced by polynomials in the differentiation operator D. See, e.g. [17, 18].

Variance is often not constant as above, but varies, perhaps randomly; ARCH (autoregressive conditionally heteroscedastic) models were introduced to apply time-series methods here, and these were generalised to GARCH models, for which, see, e.g. [38]. The general case is GARCH(p, q); GARCH(1, 1) is fairly widely applicable, and we confine ourselves to it here for simplicity.

As before, one can generalise GARCH to continuous time, COGARCH; see, e.g. [20, 57, 58]. Again, as before, COGARCH(1, 1) suffices for many purposes, so we confine ourselves to it here (see [41]).

The theory of OPUC, used in Section 1 in discrete time, finds its continuous generalisation in the theory of Krein systems [27, 36], and in particular Krein's theory of strings [31].

4. Stochastic volatility and volatility clustering

Not only does volatility vary, so that stochastic volatility models are used (as in the GARCH models of Section 3), one of the stylised facts of mathematical finance is *volatility clustering*. Economic life typically consists of periods of normality punctuated by crises; the crises give rise to clusters of high volatility. This clustering of extremes happens quite generally [2]; for background on extremes, see, e.g. [34].

In the GARCH case, stationarity requires restrictions on the parameters [74]. These involve Liapounov exponents (and Kingman's subadditive ergodic theorem). These can be calculated explicitly in the GARCH(1, 1) case, which is usually all that is needed in practice [73, Section 5.3.2]. Stationarity also brings in a stochastic recurrence equation (SRE), whose solution again involves Liapounov exponents. One obtains regular variation of the tails of the stationary process (see [73, Section 5.5.2]; this uses a result of Kesten in 1973 in the multidimensional case and of Goldie in 1991 in one dimension). A stationary GARCH process is strongly mixing and has exponentially decaying correlations, under general conditions (of Markov character; see [73, Section 5.6.1]). To recapitulate, for stationary GARCH(1, 1), one has heavy (regularly varying) tails and exponentially decreasing correlations.

This connection between the time behaviour (correlation decay) and the distribution at fixed times seems to be a reflection of the Markovian nature of the process (GARCH in this case). In the stationary case of Section 1, one can have any correlation structure and any distribution; there is no link between them, even when the process is Markovian. For, any distribution can be the stationary distribution of a Markov chain, as in the Metropolis–Hastings algorithm (see, e.g. [87]). Although the Cramér process Y encodes the time behaviour of X in (CR), the distribution of $X_0 = \int_{\mathbb{T}} \mathrm{d}Y = Y(\mathbb{T})$ (the total mass of the complex measure Y) is arbitrary. One has no way to model volatility clustering, and this restricts the suitability of the prediction theory of Section 1 for financial time series.

There is, however, no difficulty in capturing volatility clustering in other ways, in one or many dimensions; see, e.g. [13]. In finance, the dimensionality d is often high, to reflect the holding of a balanced portfolio of many risky assets, by Markowitzian diversification. The *multivariate elliptic processes* of [13] model the time evolution of such a d-dimensional time series using the theory of *elliptical distributions* [32]. One works *semi-parametrically*. The *parametric* part of the model is (Σ, μ), where Σ is the $d \times d$ covariance matrix and μ is the d-vector of means, both essential ingredients, from Markowitz' work [69, 70] (think of risk and return together; (co)variances measure risk, means measure return). The *nonparametric* part is a one-dimensional process, the *risk driver* R. Thus, the burden of dimensionality is born by (Σ, μ); R models the ambient economic or financial climate. Though this is certainly not one-dimensional in reality, one often works with a one-dimensional proxy for it, such as a stock-exchange index.

5. Higher dimensions

Much of the one- and finite-dimensional theory (for which, see, e.g. [10, 11]) extends to infinitely many dimensions, but not all of it does. The simplest infinite-dimensional setting is a (separable) Hilbert space; beyond that one has (separable) Banach space; beyond that, locally convex topological vector spaces. Our main references here are the books of Vakhania *et al.* [95] on Banach spaces, Kakihara [51], Schwartz [89] on cylindrical measures, and *The Pesi Masani Volume* [68], and the papers of Riedle [85, 86] and (with Applebaum) [6].

5.1. Hilbert space

The simplest infinite-dimensional setting is Hilbert space, and the theory here was developed by Payen in 1967 [80]. He obtained the Wold decomposition [80, Section II.3], with two components, as above. He dealt with *factorizability* (see [80, Section II.5] and below), and, hence, obtained [80, Section II.6] a three-term decomposition, one component corresponding to the singular component of the spectral measure as before, while the 'absolutely continuous component' splits into two, one part corresponding to the 'largest factorizable part', the other

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

to the residue. He studied the Szegö condition via a series decomposition into rank-one components [80, Section II.8], and also continuous time [80, Section III].

The Hilbert-space setting brings closer together the stationary and nonstationary cases (for commentary on the distinction, see [12]). The key concept here is *harmonizability* (due to Loève in 1955, and subsequently studied by Cramér, M. M. Rao, and others). A (second-order) process $X = (X_t)$ is *harmonizable* if it is the Fourier–Stieltjes transform of a random measure Y. If Y is orthogonally scattered, X is stationary and so harmonizable. There is no converse, but on a Hilbert space H, X is harmonizable if and only if it has a stationary dilation, i.e. it is the projection of a stationary process in some larger Hilbert space K. For details, see [52, 84].

The Hilbert-space methods used by Kolmogorov in 1941 to prove (KIT) are tantamount to those used by Aronszajn [7] in 1950 in his theory of RKHS. Accordingly, the result in the Hilbert-space setting is often called the *Aronszajn–Kolmogorov theorem*. See [95, Section III.1.3] for positive-definite functions (covariance kernels), [95, Section IV.4.3] for orthogonally scattered measures, and [67] for the corresponding instance of the Kolmogorov isomorphism theorem.

The sampling theorem was extended to Hilbert space by Weston in 1949 [96].

5.2. Banach spaces

As mentioned above, our main reference for the Banach case is [95], which while it does not consider prediction theory explicitly, it is nevertheless well adapted to our purposes here.

In finite-dimensional spaces, linear prediction is always possible. In infinitely many dimensions, this is no longer so. It is possible in the space $L(Y, H)$ of continuous linear operators from a Banach space Y to a Hilbert space H [22, Section 3].

The results above of Payen [80] on the Wold–Cramér concordance in Hilbert space were extended to the Banach case by Schmidt [88] and Hajduk-Chmielewska [40].

For V and W topological vector spaces, write V^* for the dual of V and $L(V, W)$ for the space of continuous operators (linear maps) from V to W. Then V has the *factorization property* if every nonnegative $A \in L(V, V^*)$ can be factorized as

$$A = T^*T,$$

where T is a continuous operator from V into some Hilbert space (depending on A); cf. [1, Section 7.3]. The factorization property is equivalent to the continuity of $v \mapsto \langle Av, v\rangle$ for some nonnegative $A \in L(V, V^*)$; this holds for Banach spaces, and for some but not all locally convex topological vector spaces [37]. The Aronszajn–Kolmogorov theorem can be extended to locally convex spaces with the factorization property [37] and so can some aspects of prediction theory (see [3]). For harmonizability, see [84].

5.3. Gaussianity and prediction

Linear prediction is always possible if one restricts attention to the Gaussian case. Even in one dimension, there is a case for doing this: there is a hierarchy of weak-dependence conditions in the Gaussian case, and another in the general case (mixing conditions; there is some economy in working with one rather than two hierarchies; for background on such weak, strong, and intermediate conditions, see, e.g. [10]). This is why some authors restrict attention to the Gaussian case throughout, as is done in the classical book by Dym and McKean [31]; Gelfand and Vilenkin [35, Section III.3.4] discussed this, but deferred to probabilistic usage by working generally.

5.4. Gaussianity and covariances

Because of this, the whole question of Gaussian measures in infinite-dimensional settings becomes unavoidable here. This rests on Schwartz's theory of Radon measures and cylindrical

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

measures [89], and in particular on the *covariance* operator (recall that a zero-mean Gaussian process is characterized by its covariance), usually written Q. The theory of covariance operators in infinitely many dimensions is considered in detail by Vakhania *et al.* [95, Section III]. They treated the Aronszajn–Kolmogorov theorem, and showed that in Hilbert space, a measure μ being strong second order (square integrable in the strong sense) is equivalent to the covariance operator Q being *nuclear* (*trace class*). Because a nuclear operator is a product of two Hilbert–Schmidt operators and covariance, being symmetric and positive, factorizes (see below), this can be expressed alternatively in the language of Hilbert–Schmidt operators. The covariance Q factorizes as

$$Q = i_Q i_Q^*,$$

where i_Q is the inclusion map from the range of Q to the Banach space X. So (Radon) Gaussian measures have nice covariances: Q is nuclear, and the above leads to the reproducing-kernel Hilbert space associated with Q [85, Section IV].

Returning to the general (non-Gaussian) case, linear prediction fails in general in infinite-dimensional Banach spaces: there exist a (non-Gaussian) random element such that linear prediction of it by Gaussian families fails, and spectral representations of stationary processes as before in the setting of $L(Y, H)$ [22, Sections 7 and 8].

5.5. Locally convex topological vector spaces

Spaces of generalized functions are typically not normable, and neither are spaces of holomorphic functions, etc. Instead, one needs to specify the topology in terms of a collection of seminorms; with separability (enough for us here) sequences suffice. Often one can use a sequence of norms and obtain a locally convex topological vector space; the prototype here is Grothendieck's nuclear spaces of 1955, for which, see [35, Section I.3].

Some aspects of prediction theory can be taken over to the locally convex case; see, e.g. [3].

5.6. Random generalized functions

If one passes from functions to generalized functions (or Schwartz distributions, though in probability theory this would overload the word distribution), random variables are replaced by random generalized functions. This was done as early as 1954 by Itô [48]. It is argued by Gelfand and Vilenkin [35, Section III.1.2, p. 243] that this setting is actually more natural and realistic:

> However, every actual measurement is accomplished by means of an apparatus which has a certain inertia … [so] the reading which the apparatus gives is not … at the instant t but rather a certain averaged value … [via] a function characterizing the apparatus.

6. Applications

We confine ourselves here to brief comments on the applications of prediction theory in the infinite-dimensional case. This forms part of the extensive area of statistical inference for stochastic processes; see, e.g. Bosq [15], Antoniadis and Sapatinas [4] (who consider El Niño), [5]. One principal field of application is *functional data analysis*, for which, see, e.g. [33, 81, 82]. Statistical analysis of data consisting of curves is well established; one application here is *gait analysis*, used to analyse the motion of walking. Another is *yield curves* in the term structure of interest rates in finance; as with anything in finance, predictive ability is of prime practical importance. The first thing to do is to represent curves economically. Wavelet expansions are often used here; an early and dramatic application was their use to digitize the FBI fingerprint bank (without which the US criminal justice system would long since have collapsed). Also used

https://doi.org/10.1239/jap/1417528475 Published online by Cambridge University Press

here are Karhunen–Loève expansions, again, in the space variable. Much of functional data analysis is concerned with independent and identically distributed data of curves. For prediction purposes, one needs to dynamicize these results, and have the expansion coefficients evolving with time. Again, wavelets are well suited to this purpose, as they allow one to handle time and space (or time and frequency) together. As always, theory and applications have much to offer each other here. We close by citing one specific application area: the prediction of electricity consumption [25].

7. Postscript on Whittaker

We mentioned J. M. Whittaker in Section 2 in connection with the sampling theorem. John Macnaghten (Jack) Whittaker (1905–1984) was Professor of Pure Mathematics at Liverpool University from 1933–53 (seconded during World War II, when he served on Montgomery's staff in the 8th Army). Last semester, the first author lectured in Liverpool, in the Whittaker Room, which has a plaque commemorating him and the sampling theorem over the door. Whittaker went on to be Vice-Chancellor of The University of Sheffield from 1953–65. There Joe Gani founded the *Journal of Applied Probability* in 1964, whose fiftieth anniversary this volume celebrates. Joe always speaks well of Whittaker as a Vice-Chancellor whose word could be relied on—not always the case.

Whittaker was the son of E. T. (Sir Edmund) Whittaker (1873–1956), who also worked on cardinal series (in 1915). They were (from 1949 to 1956) the only father-and-son pair in the Royal Society.

8. Conclusion

The above is a necessarily brief and very partial survey of an enormous field; there is much more to be said, and we will return to these matters elsewhere. We close by pointing out that this illustrates the wonderful and inexhaustably rich interplay between mathematics, probability and statistics, pure, and applied. This is at the heart of the raison d'être of the Applied Probability Trust journals, which this volume celebrates.

It is a great pleasure for both authors to contribute to this volume. It is a particular pleasure for the first author, who has known Joe Gani all his career, and who contributed to the twenty-fifth anniversary volume in 1988. He was then a new recruit to the Editorial Board, and is now retired from it after many years as a Coordinating Editor. We congratulate the Applied Probability Trust, its journals, its editors, and its staff on their first half-century, and look forward to the next one.

References

[1] Albiac, F. and Kalton, N. J. (2006). *Topics in Banach Space Theory* (Graduate Texts Math. **233**). Springer, New York.
[2] Aldous, D. (1989). *Probability Approximations via the Poisson Clumping Heuristic*. Springer, New York.
[3] Alpay, D., Timoshenko, O. and Volok, D. (2009). Carathéodory–Fejér interpolation in locally convex topological vector spaces. *Linear Algebra Appl.* **431,** 1257–1266.
[4] Antoniadis, A. and Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* **87,** 133–158.
[5] Antoniadis, A., Paparoditis, E. and Sapatinas, T. (2006). A functional wavelet-kernel approach for time series prediction. *J. R. Statist. Soc. B* **68,** 837–857.
[6] Applebaum, D. and Riedle, M. (2010). Cylindrical Lévy processes in Banach spaces. *Proc. London Math. Soc.* **101,** 697–726.
[7] Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68,** 337–404.
[8] Belyaev, Y. K. (1959). Analytic random processes. *Theory Prob. Appl.* **4,** 402–409.
[9] Bergstrom, A. R. (1990). *Continuous time Econometric Modelling*. Oxford University Press.

[10] Bingham, N. H. (2012). Szegö's theorem and its probabilistic descendants. *Prob. Surveys* **9,** 287–324.
[11] Bingham, N. H. (2012). Multivariate prediction and matrix Szegö theory. *Prob. Surveys* **9,** 325–339.
[12] Bingham, N. H. (2014). Modelling and prediction of financial time series. *Commun. Statist. Theory Meth.* **43,** 1351–1361.
[13] Bingham, N. H., Fry, J. M. and Kiesel, R. (2010). Multivariate elliptic processes. *Statist. Neerlandica* **64,** 352–366.
[14] Boas, R. P., Jr. (1972). Summation formulas and band-limited signals. *Tôhoku Math. J.* **24,** 121–125.
[15] Bosq, D. (2000). *Linear Processes in Function Spaces. Theory and Applications* (Lecture Notes Statist. **149**). Springer, New York.
[16] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008). *Time Series Analysis. Forecasting and Control*, 4th edn. John Wiley, Hoboken, NJ.
[17] Brockwell, P. J. (2001). Continuous-time ARMA processes. In *Stochastic Processes: Theory and Methods* (Handbook Statist. **19**), Elsevier, London, pp. 249–276.
[18] Brockwell, P. J. (2001). Lévy-driven CARMA processes. *Ann. Inst. Statist. Math.* **53,** 113–124.
[19] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2nd edn. Springer, New York.
[20] Brockwell, P. J., Chadraa, E. and Lindner, A. (2006). Continuous-time GARCH processes. *Ann. Appl. Prob.* **16,** 790–826.
[21] Butzer, P. L. *et al.* (2011). The sampling theorem, Poisson's summation formula, general Parseval formula, reproducing kernel formula and the Paley–Wiener theorem for bandlimited signals—their interconnections. *Appl. Anal.* **90,** 431–461.
[22] Chobanyan, S. A. and Vakhania, N. N. (1983). The linear prediction and approximation of weak second order random elements. In *Prediction Theory and Harmonic Analysis: The Pesi Masani Volume*, eds V. Mandrekar and H. Saleti, North-Holland, Amsterdam, pp. 37–60.
[23] Cramér, H. (1942). On harmonic analysis in certain function spaces. *Ark. Mat. Astr. Fys.* **28B,** 7pp.
[24] Cramér, H. and Leadbetter, M. R. (1967). *Stationary and Related Stochastic Processes*. John Wiley, New York.
[25] Cugliari, J. (2011). Prévision non paramétrique de processus à valeurs fonctionnelles. Application à la consommation d'électricité. Doctoral Thesis, Université Paris-Sud XI.
[26] De Branges, L. (1968). *Hilbert Spaces of Entire Functions*. Prentice-Hall, Englewood Cliffs, NJ.
[27] Denisov, S. A. (2006). Continuous analogs of polynomials orthogonal on the unit circle and Krein systems. *IMRS Int. Math. Res. Surv.* **2006,** 54517.
[28] Doob, J. L. (1953). *Stochastic Processes*. John Wiley, New York.
[29] Dym, H. and McKean, H. P., Jr. (1970). Application of de Branges spaces of integral functions to prediction of stationary Gaussian processes. *Illinois J. Math.* **14,** 299–343.
[30] Dym, H. and McKean, H. P. (1970). Extrapolation and interpolation of stationary Gaussian processes. *Ann. Math. Statist.* **41,** 1817–1844.
[31] Dym, H. and McKean, H. P. (1976). *Gaussian Processes, Function Theory, and the Inverse Spectral Problem*. Academic Press, New York.
[32] Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman and Hall, London.
[33] Ferraty, F. and Romain, Y. (eds) (2010). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.
[34] Finkenstädt, B. and Rootzén, H. (eds) (2004). *Extreme Values in Finance, Telecommunications, and the Environment*. Chapman & Hall/CRC, Boca Raton, FL.
[35] Gel'fand, I. M., and Vilenkin, N. Y. (1964). *Generalized Functions*, Vol. 4. Academic Press, New York.
[36] Ginovyan, M. S. and Mikaelyan, L. V. (2010). Prediction error for continuous-time stationary processes with singular spectral densities. *Acta Appl. Math.* **110,** 327–351.
[37] Gorniak, J. and Weron, A. (1980/81). Aronszajn–Kolmogorov type theorems for positive definite kernels in locally convex spaces. *Studia Math.* **69,** 235–246.
[38] Gouriéroux, C. (1997). *ARCH Models and Financial Applications*. Springer, New York.
[39] Grenander, U. and Szegö, G. (1958). *Toeplitz Forms and Their Applications*. University of California Press, Berkeley, CA.
[40] Hajduk-Chmielewska, G. (1988). The Wold-Cramér concordance problem for Banach-space-valued stationary processes. *Studia Math.* **91,** 31–43.
[41] Haug, S., Klüppelberg, C., Lindner, A. and Zapp, M. (2007). Method of moment estimation in the COGARCH(1,1) model. *Econometric J.* **10,** 320–341.
[42] Higgins, J. R. (1977). *Completeness and Basis Properties of Sets of Special Functions* (Camb. Tracts Math. **72**). Cambridge University Press.
[43] Higgins, J. R. (1985). Five short stories about the cardinal series. *Bull. Amer. Math. Soc. (N.S.)* **12,** 45–89.
[44] Higgins, J. R. (1996). *Sampling Theory in Fourier and Signal Analysis: Foundations*. Clarendon Press, Oxford.

[45] Higgins, J. R. and Stens, R. L. (eds) (1996). *Sampling Theory in Fourier and Signal Analysis: Advanced Topics*. Oxford University Press.
[46] Inoue, A. (2008). AR and MA representations of partial autocorrelation functions, with applications. *Prob. Theory Relat. Fields* **140,** 523–551.
[47] Inoue, A. and Kasahara, Y. (2006). Explicit representation of finite predictor coefficients and its applications. *Ann. Statist.* **34,** 973–993.
[48] Itô, K. (1954). Stationary random distributions. *Mem. Coll. Sci. Univ. Kyoto Ser. A. Math.* **28**, 209–223.
[49] Janson, S. (1997). *Gaussian Hilbert Spaces* (Camb. Tracts Math. **129**). Cambridge University Press.
[50] Kabanov, Y., Liptser, R. and Stoyanov, J. (eds) (2006). *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*. Springer, Berlin.
[51] Kakihara, Y. (1997). *Multidimensional Second Order Stochastic Processes*. World Scientific, River Edge, NJ.
[52] Kakihara, Y. (1997). Dilations of Hilbert–Schmidt operator-valued measures and applications. In *Stochastic Processes and Functional Analysis: In Celebration of M. M. Rao's 65th Birthday*, eds J. A. Goldstein, N. E. Gretsky and J. J. Uhl, Dekker, New York, pp. 123–135.
[53] Kallenberg, O. (2002). *Foundations of Modern Probability*, 2nd edn. Springer, New York.
[54] Karhunen, K. (1950). Uber die Struktur stationärer zufälliger funktionen. *Ark. Mat.* **1,** 141–160.
[55] Kasahara, Y. and Bingham, N. H. (2014). Verblunsky coefficients and Nehari sequences. *Trans. Amer. Math. Soc.* **366,** 1363–1378.
[56] Katznelson, Y. (2004). *An Introduction to Harmonic Analysis*, 3rd edn. Cambridge University Press.
[57] Klüppelberg, C., Lindner, A. and Maller, R. (2004). A continuous-time GARCH process driven by a Lévy process: stationarity and second-order behaviour. *J. Appl. Prob.* **41,** 601–622.
[58] Klüppelberg, C., Lindner, A. and Maller, R. A. (2006). Continuous-time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models. In *From Stochastic Calculus to Mathematical Finance: The Shiryaev Festschrift*, eds Y. Kabanov *et al.*, Springer, Berlin, pp. 393–419.
[59] Kolmogorov, A. N. (1941). Stationary sequences in Hilbert space. *Bull. Moskov. Gos. Univ. Mat.* **2,** 1–40 (in Russian).
[60] Kolmogorov, A. N. (1986). *Selected Works of A. N. Kolmogorov*, Vol. 2. Nauka, Moskva (in Russian).
[61] Kramer, H. P. (1959). A generalised sampling theorem. *J. Math. Phys.* **38,** 68–72.
[62] Levinson, N. and McKean, H. P., Jr. (1964). Weighted trigonometrical approximation on R^1 with application to the germ field of a stationary Gaussian noise. *Acta Math.* **112,** 99–143.
[63] Lévy, P. (1948). *Processus Stochastiques et Mouvement Brownien*. Gauthier-Villars, Paris.
[64] Lloyd, S. P. (1959). A sampling theorem for stationary (wide sense) stochastic processes. *Trans. Amer. Math. Soc.* **92,** 1–12.
[65] Loève, M. (1948). Fonctions aléatoires du second ordre. Supplement to *Processus Stochastiques et Mouvement Brownien*, Gauthier-Villars, Paris, pp. 228–352.
[66] Loève, M. (1978). *Probability Theory. II* (Graduate Texts Math. **46**), 4th edn. Springer, New York.
[67] Mandrekar, V. and Salehi, H. (1972). The square-integrability of operator-valued functions with respect to a non-negative operator-valued measure and the Kolmogorov isomorphism theorem. *Indiana Univ. Math. J.* **20,** 545–563.
[68] Mandrekar, V. and Salehi, H. (eds) (1983). *Prediction Theory and Harmonic Analysis: The Pesi Masani Volume*. North-Holland, Amsterdam.
[69] Markowitz, H. (1952). Portfolio selection. *J. Finance* **7,** 77–91.
[70] Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments*. John Wiley, New York.
[71] Martin, R. T. W. (2010). Symmetric operators and reproducing kernel Hilbert spaces. *Complex Anal. Operat. Theory* **4,** 845–880.
[72] Masani, P. (1968). Orthogonally scattered measures. *Adv. Math.* **2,** 61–117.
[73] Mikosch, T. (2000). Modelling dependence and tails of financial time series. In *Symposium in Honour of Ole E. Barndorff-Nielsen*, Aarhus University, pp. 61–73.
[74] Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6,** 318–334.
[75] Niemi, H. (1975). Stochastic processes as Fourier transforms of stochastic measures. *Ann. Acad. Sci. Fenn. Ser. A I* **591,** 47pp.
[76] Nikol'skiĭ, N. K. (1986). *Treatise on the Shift Operator. Spectral Function Theory* (Grundl. Math. Wiss. **273**). Springer, Berlin.
[77] Nikolski, N. K. (2002). *Operators, Functions and Systems: An Easy Reading*, Vol. 1. American Mathematical Society, Providence, RI.
[78] Paley, R. E. A. C. and Wiener, N. (1934). *Fourier Transforms in the Complex Domain*. American Mathematical Society, New York.
[79] Partington, J. R. (1997). *Interpolation, Identification, and Sampling*. Oxford University Press.
[80] Payen, R. (1967). Fonctions aléatoires du second ordre à valeurs dans un espace de Hilbert. *Ann. Inst. H. Poincaré B (N.S.)* **3,** 323–396.

[81] Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.
[82] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.
[83] Rao, M. M. (1985). Harmonizable, Cramér, and Karhunen classes of processes. In *Time Series in the Time Domain* (Handbook Statist. **5**), North-Holland, Amsterdam, pp. 279–310.
[84] Richard, P. H. (1992). Harmonizability, V-boundedness and stationary dilation for Banach-valued processes. In *Probability in Banach Spaces*, eds R. M. Dudley, M. G. Hahn and J. Kuelbs, Birkhäuser, Basel, pp. 189–205.
[85] Riedle, M. (2011). Cylindrical Wiener processes. In *Séminaire de Probabilités XLIII* (Lecture Notes Math. **2006**), Springer, Berlin, pp. 191–214.
[86] Riedle, M. (2011). Infinitely divisible cylindricalmeasures on Banach spaces. *Studia Math.* **207,** 235–256.
[87] Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer, New York.
[88] Schmidt, F. (1978). Banach-space-valued stationary processes with absolutely continuous spectral function. In *Probability Theory on Vector Spaces* (Lecture Notes Math. **656**), Springer, Berlin, pp. 237–244.
[89] Schwartz, L. (1973). *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*. Oxford University Press.
[90] Shiryaev, A. N. (1989). Kolmogorov: life and creative activities. *Ann. Prob.* **17,** 866–944.
[91] Simon, B. (2005). *Orthogonal Polynomials on the Unit Circle. Part 1. Classical Theory*. American Mathematical Society, Providence, RI.
[92] Simon, B. (2005). *Orthogonal Polynomials on the Unit Circle. Part 2. Spectral Theory*. American Mathematical Society, Providence, RI.
[93] Simon, B. (2011). *Szegő's Theorem and its Descendants. Spectral Theory for L^2 Perturbations of Orthogonal Polynomials*. Princeton University Press.
[94] Szegö, G. (1939). *Orthogonal Polynomials*. American Mathematical Society, New York.
[95] Vakhania, N. N., Tarieladze, V. I. and Chobanyan, S. A. (1987). *Probability Distributions on Banach Spaces*. Reidel, Dordrecht.
[96] Weston, J. D. (1949). The cardinal series in Hilbert space. *Proc. Camb. Phil. Soc.* **45,** 335–341.
[97] Whittaker, J. M. (1935). *Interpolatory Function Theory* (Camb. Tracts Math. **33**). Cambridge University Press.
[98] Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series. With Engineering Applications*. John Wiley, New York.
[99] Wiener, N. (1981). *Collected Works With Commentaries: The Hopf–Wiener Integral Equation; Prediction and Filtering; Quantum Mechanics and Relativity; Miscellaneous Mathematical Papers* (ed. P. R. Masani), Vol. III. MIT Press, Cambridge, MA.

N. H. BINGHAM, *Imperial College London*

Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
Email address: n.bingham@imperial.ac.uk

BADR MISSAOUI, *Imperial College London*

Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.
Email address: badr.missaoui08@imperial.ac.uk