Cointegrated linear processes in Hilbert space

Brendan K. Beare[1] and Juwon Seo[2]

[1]Department of Economics, University of California, San Diego
[2]Department of Economics, National University of Singapore

November 30, 2015

Abstract

Consider a sequence $X = (X_0, X_1, \ldots)$ of random elements of a real separable Hilbert space H, whose first difference is a standard linear process with long run covariance operator Λ. We show that, under a mild summability condition, the set of vectors $x \in H$ for which $\langle X_t, x\rangle$ is stationary given a suitable initialization X_0 is the kernel of Λ. We call this space the cointegrating space, and its orthogonal complement the attractor space. Our main result is a version of the Granger representation theorem: we show that if X follows an autoregressive law of motion, then the attractor space is the kernel of the impact operator in the error correction representation of X, and the cointegrating space is the closure of the range of the adjoint of the impact operator. Our proof appears to be novel even when specialized to the case where H is finite dimensional Euclidean space.

We thank Juan Carlos Escanciano and Werner Ploberger for useful suggestions at an early stage, and Jim Hamilton, Peter Hansen, Lajos Horváth and Peter Phillips for helpful discussions. Much of this work was completed while the first author was visiting the National University of Singapore during September 2015.

1 Introduction

The subject of time series analysis has traditionally dealt with time series that take values in finite dimensional Euclidean space $\mathbb{R}^n$. A more recent literature on so-called functional time series analysis deals with time series that take values in a possibly infinite dimensional Banach or Hilbert space, frequently a space of functions. For instance, one observation of such a time series may be a continuous record of the value of some asset over a given trading day, or the distribution of income across an economy in a given year. An important early contribution to the literature on functional time series is the monograph of Bosq (2000), which gives a detailed theoretical treatment of linear processes in Banach and Hilbert spaces. Hörmann and Kokoszka (2012) and Horváth and Kokoszka (2012) discuss much of the subsequent literature. Empirical applications of functional time series analysis to economic and financial data have studied the term structure of interest rates (Kargin and Onatski, 2008), intraday cumulative returns (Kokoszka and Zhang, 2012), intraday volatility (Hörmann et al., 2013; Gabrys et al., 2013), and the distributions of high frequency stock returns and of individual earnings (Park and Qian, 2012; Chang et al., 2015).

The property of cointegration, well studied for time series in finite dimensional Euclidean space, was first introduced by Granger (1981). Its study transformed the practice of time series econometrics over the following two decades, especially in applications to macroeconomic data. Given the recent surge of interest in functional data analysis and functional time series, an extension of the methods of cointegration analysis to a functional time series setting may be valuable. A recent paper by Chang et al. (2015) appears to be the first effort in this direction. Those authors consider a time series taking values in a Hilbert space of square integrable centered probability density functions. They introduce a notion of cointegration for this space, and develop associated statistical methods based on functional principal components analysis.

In this paper we build on the contribution of Chang et al. (2015) by taking a closer look at what it means for a functional time series to be cointegrated. We show that the cointegrating space may be sensibly defined as the kernel of the long run covariance operator associated with the difference of our time series, while the attractor space (which Chang et al. call the unit root space) may be sensibly defined as its orthogonal complement. Our main result is a version of the Granger representation theorem: we show that if the levels of our time series follow an autoregressive law of motion, then the attractor space is the kernel of the impact operator in the error correction representation, and the cointegrating space is the closure of the range of the adjoint of the impact operator. Our proof appears to be novel even when specialized to the case where our time series takes values in finite dimensional Euclidean space.

To focus on the essential aspects of cointegration in a general Hilbert space, we make a number of simplifying assumptions. In particular, consistent with much of the literature on functional time series, we only consider processes that are purely stochastic, with no deterministic component.

Our paper is organized as follows. In Section 2 we review some essential mathematics. In Section 3 we discuss what it means for a linear process in Hilbert space to be cointegrated and provide some related results. Our main result, on the Granger representation theorem, is given in Section 4.

2 Essential preliminaries

Here we briefly review essential background material for the study of cointegrated linear processes in Hilbert space, and fix standard notation and terminology. A reader familiar with the literature on functional time series analysis can likely skip this section. Our primary sources are Conway (1990) and Bosq (2000).

2.1 Continuous linear operators on Hilbert space

Let H be a real separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We denote by $\mathcal{L}_H$ the space of continuous linear operators from H to H equipped with the operator norm

$$\|A\|_{\mathcal{L}_H} = \sup_{\|x\| \leq 1} \|A(x)\|. \tag{2.1}$$

It can be shown that $\mathcal{L}_H$ is a Banach space. We denote the kernel of an operator $A \in \mathcal{L}_H$ by

$$\ker A = \{x \in H : A(x) = 0\}, \tag{2.2}$$

and its range by

$$\operatorname{ran} A = \{A(x) : x \in H\}. \tag{2.3}$$

Both are linear subspaces of H. The dimension of $\ker A$ is called the nullity of A, and the dimension of $\operatorname{ran} A$ is called the rank of A. The rank-nullity theorem asserts that the two must sum to the dimension of H.

To each $A \in \mathcal{L}_H$ there corresponds an operator $A^* \in \mathcal{L}_H$, called the adjoint of A, that is uniquely defined by the property

$$\langle A(x), y\rangle = \langle x, A^*(y)\rangle \text{ for all } x, y \in H. \tag{2.4}$$

Note that $A = A^{**}$. If $A = A^*$, we say that A is self-adjoint.

An operator $A \in \mathcal{L}_H$ is said to be positive semidefinite if

$$\langle A(x), x\rangle \geq 0 \text{ for all } x \in H, \tag{2.5}$$

and positive definite if the inequality is strict for all nonzero $x \in H$. It is said to be compact if it is the limit of a sequence of finite rank operators in $\mathcal{L}_H$.

Given a set $G \subseteq H$, we denote the orthogonal complement to G by

$$G^\perp = \{x \in H : \langle x, y\rangle = 0 \text{ for all } y \in G\}, \tag{2.6}$$

and the closure of G—that is, the union of G and its limit points—by $\mathrm{cl}(G)$. We will make repeated use of the following equalities, valid for any $A \in \mathcal{L}_H$ (see e.g. Conway, 1990, pp. 35–36):

$$\ker A = (\operatorname{ran} A^*)^\perp \quad \text{and} \quad (\ker A)^\perp = \mathrm{cl}(\operatorname{ran} A^*). \tag{2.7}$$

2.2 Random elements of Hilbert space

Let $(\Omega, \mathcal{F}, P)$ be our underlying probability space. A random element of H is a measurable map $Z : \Omega \to H$, where H is understood to be equipped with its Borel σ-algebra. Noting that $\|Z\|$ is a real valued random variable, if $E\|Z\| < \infty$ we say that Z is integrable. If Z is integrable then there exists a unique element of H, which we denote EZ, such that

$$E\langle Z, x\rangle = \langle EZ, x\rangle \text{ for all } x \in H. \tag{2.8}$$

We call EZ the expected value of Z. It is also called the Bochner integral of Z.

Let L^2_H be the space of random elements Z of H (identifying random elements that are equal almost surely) that satisfy $E\|Z\|^2 < \infty$ and $EZ = 0$, equipped with the norm

$$\|Z\|_{L^2_H} = (E\|Z\|^2)^{1/2}, \quad Z \in L^2_H. \tag{2.9}$$

It can be shown that L^2_H is a Banach space. For $Z \in L^2_H$, the Cauchy-Schwarz inequality implies that $\langle Z, x\rangle Z$ is integrable for each $x \in H$. We may therefore define the operator $C_Z \in \mathcal{L}_H$ by

$$C_Z(x) = E\left(\langle Z, x\rangle Z\right), \quad x \in H. \tag{2.10}$$

We call C_Z the covariance operator of Z. It can be shown that C_Z is self-adjoint, positive semidefinite, and compact (and in fact nuclear; see Bosq, 2000, p. 37).

2.3 Linear processes in Hilbert space

Let $\eta = (\eta_t, t \in \mathbb{Z})$ be an independent and identically distributed (iid) sequence in L^2_H, and let $(A_k, k \geq 0)$ be a sequence in $\mathcal{L}_H$ satisfying $\sum_{k=0}^{\infty} \|A_k\|^2_{\mathcal{L}_H} < \infty$. Then it can be shown (Bosq, 2000, p. 182) that for each $t \in \mathbb{Z}$ the series

$$Z_t = \sum_{k=0}^{\infty} A_k(\eta_{t-k}) \tag{2.11}$$

is convergent in L^2_H. We call the sequence $(Z_t, t \in \mathbb{Z})$ a linear process in H with innovations η. More generally, given any $t_0 \in \mathbb{Z} \cup \{-\infty\}$, we call the sequence $(Z_t, t > t_0)$ a linear process in H with innovations η. Such a linear process is necessarily stationary.

When the operators in (2.11) satisfy $\sum_{k=0}^{\infty} \|A_k\|_{\mathcal{L}_H} < \infty$, our linear process $(Z_t, t > t_0)$ is said to be a standard linear process in H. In this case the series $A = \sum_{k=0}^{\infty} A_k$ is convergent in $\mathcal{L}_H$, and we define the long run covariance operator $V \in \mathcal{L}_H$ for our standard linear process to be the composition

$$V = AC_{\eta_0}A^*, \tag{2.12}$$

where $C_{\eta_0} \in \mathcal{L}_H$ is the covariance operator of η_0. Note that V is simply the covariance operator of $A(\eta_0)$, and is therefore self-adjoint, positive semidefinite and compact (indeed nuclear).

3 Cointegration in Hilbert space

3.1 Basic setup

Our time series of interest $X = (X_t, t \geq 0)$ is a sequence in L^2_H. Denote the first difference of X by $\Delta X = (\Delta X_t, t \geq 1)$, with $\Delta X_t = X_t - X_{t-1}$. Leaving aside the choice of the initial condition $X_0 \in L^2_H$, we suppose that ΔX is a standard linear process in H. It therefore admits the representation

$$\Delta X_t = \sum_{k=0}^{\infty} \Psi_k(\varepsilon_{t-k}), \quad t \geq 1, \tag{3.1}$$

where $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ is an iid sequence in L^2_H, and $(\Psi_k, k \geq 0)$ is a sequence in $\mathcal{L}_H$ satisfying $\sum_{k=0}^{\infty} \|\Psi_k\|_{\mathcal{L}_H} < \infty$. As discussed in Section 2.3, this summability condition is sufficient for the series in (3.1) to be convergent in L^2_H and for the series $\Psi = \sum_{k=0}^{\infty} \Psi_k$ to be convergent in $\mathcal{L}_H$, and ΔX is necessarily stationary. Nonetheless, for reasons to become apparent we require that the stronger condition

$$\sum_{k=0}^{\infty} k\|\Psi_k\|_{\mathcal{L}_H} < \infty \tag{3.2}$$

is satisfied. We require the covariance operator Σ of ε_0 to be positive definite and denote the long run covariance operator of ΔX by $\Lambda = \Psi\Sigma\Psi^*$, as in (2.12) above. Note that since Σ is positive definite, it must be the case that

$$\ker \Lambda = \ker \Psi^*. \tag{3.3}$$

Assumption 3.1. The differenced process ΔX satisfies (3.1) for some iid sequence $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ in L^2_H and some sequence $(\Psi_k, k \geq 0)$ in $\mathcal{L}_H$ satisfying (3.2). The covariance operator Σ of ε_0 is positive definite.

3.2 Beveridge-Nelson decomposition

The purpose of condition (3.2) is to facilitate a version of the Beveridge-Nelson decomposition for X. For $k \geq 0$ define $\tilde{\Psi}_k = -\sum_{j=k+1}^{\infty} \Psi_j$. Under (3.2) we have

$$\sum_{k=0}^{\infty} \|\tilde{\Psi}_k\|_{\mathcal{L}_H} \leq \sum_{k=0}^{\infty} \sum_{j=k+1}^{\infty} \|\Psi_j\|_{\mathcal{L}_H} = \sum_{k=0}^{\infty} k\|\Psi_k\|_{\mathcal{L}_H} < \infty. \tag{3.4}$$

Recalling our discussion in Section 2.3, condition (3.4) ensures that the series $\nu_t = \sum_{k=0}^{\infty} \tilde{\Psi}_k \varepsilon_{t-k}$ is convergent in L^2_H for each $t \in \mathbb{Z}$, and that the sequence $\nu = (\nu_t, t \geq 0)$ in L^2_H is stationary. For $t \geq 1$ let $\xi_t = \sum_{s=1}^{t} \varepsilon_s$, and let $\xi_0 = 0 \in L^2_H$. With a little algebra we obtain

$$X_t = X_0 - \nu_0 + \Psi(\xi_t) + \nu_t, \quad t \geq 0. \tag{3.5}$$

The decomposition (3.5) was introduced by Beveridge and Nelson (1981) for the case $H = \mathbb{R}$ to study macroeconomic business cycle fluctuations, and is discussed in detail by Phillips and Solo (1992). For the multidimensional case $H = \mathbb{R}^n$ it is often known as the common trends representation of Stock and Watson (1988). It was first applied in an infinite dimensional Hilbert space setting by Chang et al. (2015). The point is that X_t is decomposed into the sum of an initial condition $X_0 - \nu_0$, a random walk component $\Psi(\xi_t)$, and a transitory component ν_t.

3.3 Cointegrating and attractor spaces

Definition 3.1. We will call the kernel of Λ the *cointegrating space*, and its orthogonal complement the *attractor space*.

Our next two results explain why the terminology introduced in Definition 3.1 is sensible. In Proposition 3.1 we show that the cointegrating space consists precisely of those vectors whose inner products with the levels of X are stationary given suitable initialization, consistent with the use of the term in prior literature. In Proposition 3.2 we show that the random walk component $\Psi(\xi_t)$ in the Beveridge-Nelson decomposition (3.5) is confined to the attractor space. Given a suitable initialization, deviations of our process from the attractor space may therefore be attributed to the transitory component ν_t, making protracted large deviations unlikely. The term attractor is taken from Johansen (1995, p. 41).

Proposition 3.1. *Under Assumption 3.1, an element $x \in H$ belongs to the cointegrating space if and only if, for some choice of the initial condition $X_0 \in L^2_H$, the sequence of real valued random variables $X^{(x)} = (\langle X_t, x\rangle, t \geq 0)$ is stationary. Moreover, a single choice of the initial condition can make $X^{(x)}$ stationary for all x in the cointegrating space.*

Proof. Taking the inner product of both sides of (3.5) with $x \in H$ gives

$$\langle X_t, x\rangle = \langle X_0 - \nu_0, x\rangle + \langle \Psi(\xi_t), x\rangle + \langle \nu_t, x\rangle, \quad t \geq 0. \tag{3.6}$$

Since $\Psi(\xi_t)$ has covariance operator $t\Psi\Sigma\Psi^* = t\Lambda$, the term $\langle\Psi(\xi_t), x\rangle$ has second moment $t\langle\Lambda(x), x\rangle$. Therefore if $x \in \ker\Lambda$ then $\langle\Psi(\xi_t), x\rangle$ vanishes and, employing the initialization $X_0 = \nu_0$, we obtain $X^{(x)} = \nu^{(x)}$, where we denote $\nu^{(x)} = (\langle \nu_t, x\rangle, t \geq 0)$. We know that $\nu^{(x)}$ is stationary because ν is stationary and $\langle\cdot, x\rangle$ is linear, hence Borel measurable. Thus $X^{(x)}$ is stationary.

Suppose instead that $x \notin \ker\Lambda$. Then $x \notin \ker\Psi^*$ by (3.3), and so $\langle\Lambda(x), x\rangle = \langle\Sigma\Psi^*(x), \Psi^*(x)\rangle > 0$ due to the positive definiteness of Σ. Thus $E\langle\Psi(\xi_t), x\rangle^2 = t\langle\Lambda(x), x\rangle$ is increasing in t. Using (3.6) and Minkowski's inequality we can bound the square root of $E\langle\Psi(\xi_t), x\rangle^2$ by the quantity

$$(E\langle X_t, x\rangle^2)^{1/2} + (E\langle \nu_t, x\rangle^2)^{1/2} + (E\langle X_0 - \nu_0, x\rangle^2)^{1/2}, \tag{3.7}$$

which is finite for any $X_0 \in L^2_H$. Since $\nu^{(x)}$ is stationary, $E\langle\nu_t, x\rangle^2$ cannot depend on t, and so the fact that $E\langle\Psi(\xi_t), x\rangle^2$ increases with t implies that $E\langle X_t, x\rangle^2$ must also increase with t. Thus $X^{(x)}$ cannot be stationary.

Since the initialization $X_0 = \nu_0$ used to make $X^{(x)}$ stationary for $x \in \ker\Lambda$ did not depend on x, the final assertion of the Proposition is also proved. □

Proposition 3.2. *Under Assumption 3.1, the attractor space is equal to the closure of the range of* Ψ.

Proof. Immediate from (2.7) and (3.3). $\square$

Remark 3.1. In their discussion of cointegration in a Hilbert space setting, Chang et al. (2015) assume that H can be written as the direct sum of two subspaces H_N and H_S, such that $(\langle X_t, x\rangle, t \geq 0)$ is nonstationary for all $x \in H_N$ and stationary for all $x \in H_S$. Proposition 3.1 shows that this is always possible if we simply take $H_S = \ker \Lambda$ and $H_N = (\ker \Lambda)^{\perp}$ and suitably initialize X. In addition, Proposition 3.2 ensures that Assumption 2.1(b) of Chang et al. is always satisfied, since it implies that $\operatorname{ran} \Psi \subseteq H_N$ and that, when H_N is finite dimensional, $\operatorname{ran} \Psi = H_N$.

3.4 More on the initial condition

In the proof of Proposition 3.1 we observed that the initialization $X_0 = \nu_0$ makes $X^{(x)}$ stationary for all x in the cointegrating space. There are other initializations that can be used to achieve the same effect.

Proposition 3.3. *Under Assumption 3.1, if* $X_0 = \nu_0 + \zeta$ *for some* $\zeta \in L^2_H$ *independent of the innovation sequence* ε*, then an element* $x \in H$ *is in the cointegrating space if and only if* $X^{(x)}$ *is stationary.*

Proof. If $x \notin \ker \Lambda$ then Proposition 3.1 implies that $X^{(x)}$ is nonstationary. If $x \in \ker \Lambda$ and if X_0 is of the stated form then from (3.3) and (3.6) we have that

$$\langle X_t, x\rangle = \langle \zeta, x\rangle + \langle \nu_t, x\rangle, \quad t \geq 0. \tag{3.8}$$

Independence of ζ and ε implies independence of $\langle \zeta, x\rangle$ and $\nu^{(x)}$. Since $\nu^{(x)}$ is stationary, it follows that $X^{(x)}$ is stationary. $\square$

A caveat on Proposition 3.3 is perhaps in order. Though an initial condition of the form $X_0 = \nu_0 + \zeta$ with $\zeta \in L^2_H$ independent of ε makes $X^{(x)}$ stationary for all $x \in \ker \Lambda$, it does not ensure that $X^{(x)}$ is ergodic. Applying Theorem 7.7 of Bosq (2000), for every $x \in \ker \Lambda$ we have $n^{-1} \sum_{t=1}^{n} \langle X_t, x\rangle \to \langle \zeta, x\rangle$ as $n \to \infty$ almost surely. If we would like $X^{(x)}$ to be not merely stationary but also ergodic for every $x \in \ker \Lambda$, we should therefore choose ζ to satisfy $P(\zeta \in (\ker \Lambda)^{\perp}) = 1$. With this choice of ζ, for all $x \in \ker \Lambda$ we have $n^{-1} \sum_{t=1}^{n} \langle X_t, x\rangle \to 0 = E\langle X_0, x\rangle$ as $n \to \infty$ almost surely, meaning that $X^{(x)}$ is stationary and ergodic for all $x \in \ker \Lambda$.

3.5 Orders of integration

In time series analysis it is common to say that a series is integrated of order d or $I(d)$, frequently $I(0)$ or $I(1)$. As discussed by Davidson (2009), the precise meaning of this expression varies by author and context, and is not always explicitly given. Following the approach of Johansen (1995, p. 35), we will say that a sequence in L^2_H is $I(0)$ if it is a standard linear process with nonzero long run covariance operator, and that a sequence in L^2_H is $I(d)$ with $d = 1, 2, \ldots$ if its d^{th} difference is $I(0)$. Note that the difference of a standard linear process is a standard linear process with long run covariance operator equal to zero.

In this paper we are primarily concerned with the case where X is $I(1)$. Assumption 3.1 asserts that the first difference of X is a standard linear process, implying that X is $I(1)$ if and only if $\Lambda \neq 0$. Moreover, Assumption 3.1 rules out the possibility that X is $I(d)$ for any $d \geq 2$, because if $\Lambda = 0$ then the d^{th} difference of X is a standard linear process with long run covariance equal to zero for every $d \geq 0$. In the degenerate case $\Lambda = 0$ it is not necessarily the case that X suitably initialized is $I(0)$, because our assumptions do not guarantee that X will itself have nonzero long run covariance operator. We may nevertheless assert that X suitably initialized is stationary.

Proposition 3.4. *Under Assumption 3.1, X is stationary for some choice of the initial condition $X_0 \in L^2_H$ if and only if $\Lambda = 0$.*

Proof. It follows from (3.3) that $\Lambda = 0$ if and only if $\Psi = 0$. If $\Psi = 0$ then we see from (3.5) that X is stationary given the initialization $X_0 = \nu_0$. If $\Psi \neq 0$ then we may choose an $x \in H$ with $x \notin \ker \Psi^*$. Taking the inner product of both sides of (3.5) with x and arguing as in the proof of Proposition 3.1, we find that $E\langle X_t, x\rangle^2$ is increasing in t for any choice of $X_0 \in L^2_H$, ruling out stationarity of X. □

4 A Hilbert space version of Granger's theorem

We now suppose that the process X defined in the previous section satisfies an autoregressive law of motion.

Assumption 4.1. The process X satisfies Assumption 3.1, and further, we have

$$X_t = \sum_{j=1}^{p} \Phi_j(X_{t-j}) + \varepsilon_t, \quad t \geq p, \tag{4.1}$$

for some $p \in \mathbb{N}$ and some $\Phi_1, \ldots, \Phi_p \in \mathcal{L}_H$.

Let $\Phi = \sum_{j=1}^{p} \Phi_j$ and $\Pi = \Phi - \mathrm{id}_H$, where id_H is the identity operator on H. Subtracting X_{t-1} from both sides of (4.1), we obtain

$$\Delta X_t = \Pi(X_{t-1}) + \sum_{j=1}^{p-1} \tilde{\Phi}_j \Delta X_{t-j} + \varepsilon_t, \quad t \geq p, \tag{4.2}$$

where we have defined $\tilde{\Phi}_j = -\sum_{i=j+1}^{p} \Phi_i$, and the sum over j should be interpreted as zero when $p = 1$. Equation (4.2) is known as the error correction representation of X. We will refer to Π as the impact operator.

The Granger representation theorem, also known simply as Granger's theorem, has a complicated history. It first appeared in published form in a paper of Engle and Granger (1987). It is labeled there as the "Granger representation theorem", owing to the fact that it had appeared in an earlier unpublished working paper of Granger (1983). Granger's theorem relates the cointegrating and attractor spaces to the impact operator Π, which in the case $H = \mathbb{R}^n$ considered by Granger can be viewed as an $n \times n$ matrix. Its central assertion is that Π may be factorized as $\Pi = \alpha\beta'$, where α and β are full rank $n \times r$ matrices, the columns of β span the cointegrating space, and the prime indicates matrix transposition. Unfortunately the original proof of Granger's theorem contained an error: a counterexample eventually provided by Johansen (2009, p. 126) shows that Lemma 1 of Engle and Granger (1987), which is also Theorem 1 of Granger (1983), is incorrect. This flaw made it unclear whether the matrix α was of full rank. In the account of Granger's theorem given in the well known textbook by Hamilton (1994), the statement of the theorem is in fact weakened so that α is not claimed to be of full rank. Through a series of papers on Granger's theorem by Johansen (1988, 1991, 1992), eventually culminating in a monograph (Johansen, 1995), it became clear that we can ensure that α is full rank by ruling out solutions to (4.1) that are $I(d)$ for $d > 1$. Engle and Granger (1987) did in fact assume that their solution to (4.1) was $I(1)$, and here our Assumption 3.1 rules out $I(d)$ solutions with $d > 1$.

The main result of our paper is a version of Granger's theorem that applies in a general real separable Hilbert space. The proof is, as far as we can tell, novel even when specialized to the classical case $H = \mathbb{R}^n$. Some of the arguments used by Johansen (1995) do not readily extend to an infinite dimensional setting; in particular, we do not have a well defined notion of the determinant of an operator in $\mathcal{L}_H$, and it can be difficult to establish that an operator is invertible.

Theorem 4.1. *Suppose that Assumption 4.1 holds and that the kernel of* Π *is finite dimensional. Then the attractor space is the kernel of* Π *and the cointegrating space is the closure of the range of* Π^**.*

Remark 4.1. The first and second parts of the conclusion of Theorem 4.1 are of course equivalent in view of (2.7). When $H = \mathbb{R}^n$, either part is equivalent to the assertion that the $n \times n$ matrix Π may be factorized as $\Pi = \alpha\beta'$, where α and β are full rank $n \times r$ matrices, and the columns of β span the cointegrating space. To see this, we let $\bar{\beta}$ denote an $n \times n$ matrix whose first r columns form a basis for the cointegrating space and whose last $n - r$ columns form a basis for the attractor space, and we let $\bar{\alpha}' = \bar{\beta}^{-1}\Pi'$, so that $\Pi = \bar{\alpha}\bar{\beta}'$. By construction,

$$\{\bar{\beta}'x : x \in (\ker \Lambda)^{\perp}\} = \{x \in \mathbb{R}^n : x_1 = \cdots = x_r = 0\}. \tag{4.3}$$

We thus find that $\ker \Pi = (\ker \Lambda)^{\perp}$ if and only if the first r columns of $\bar{\alpha}$ are linearly independent and the last $n - r$ columns of $\bar{\alpha}$ are zero. In this case we may set α and β equal to the first r columns of $\bar{\alpha}$ and $\bar{\beta}$ respectively.

Remark 4.2. Taken together, Proposition 3.2 and Theorem 4.1 imply that the kernel of Π is the closure of the range of Ψ, and that the kernel of Ψ^* is the closure of the range of Π^*. This is the "interesting symmetry" to which Johansen (1991, p. 1561) refers in his discussion of Granger's theorem.

Remark 4.3. If the autoregressive operators $\Phi_1, \ldots, \Phi_p$ are compact, then the impact operator Π is a Fredholm operator, and must therefore have finite dimensional kernel (Conway, 1990, pp. 349–350).

Our proof of Theorem 4.1 relies on five lemmas to be stated.

Lemma 4.1. *Under Assumption 4.1 we have* $(\ker \Lambda)^{\perp} \subseteq \ker \Pi$.

Proof. Since ΔX is a stationary linear process with innovations ε, it is clear from (4.2) that $(\Pi(X_t), t \geq 0)$ is stationary. Thus $(\langle X_t, \Pi^*(x)\rangle, t \geq 0)$ must also be stationary for each $x \in H$. It now follows from Proposition 3.1 that $\operatorname{ran} \Pi^* \subseteq \ker \Lambda$, and hence $(\ker \Lambda)^{\perp} \subseteq (\operatorname{ran} \Pi^*)^{\perp}$. But $(\operatorname{ran} \Pi^*)^{\perp} = \ker \Pi$ by (2.7), so we conclude that $(\ker \Lambda)^{\perp} \subseteq \ker \Pi$. □

Remark 4.4. In the classical case $H = \mathbb{R}^n$, Lemma 4.1 is, essentially, the weaker version of Granger's theorem stated as Proposition 19.1 in the textbook of Hamilton (1994). The inclusion $(\ker \Lambda)^{\perp} \subseteq \ker \Pi$ is equivalent to the existence of a factorization $\Pi = \alpha\beta'$, where β is a full rank $n \times r$ matrix whose columns span the cointegrating space, and α is an $n \times r$ matrix that may or may not be of full rank. This is a simple result; the inclusion $\ker \Pi \subseteq (\ker \Lambda)^{\perp}$, which together with Lemma 4.1 establishes Theorem 4.1, is significantly more difficult to show.

Lemma 4.2. *Under Assumption 4.1 we have* $\Pi(X_0 - \nu_0) = 0$.

Proof. Since $(\ker \Lambda)^{\perp} \subseteq \ker \Pi$ by Lemma 4.1, we have from Proposition 3.2 that $\operatorname{ran} \Psi \subseteq \ker \Pi$. Thus, combining the error correction representation (4.2) and Beveridge-Nelson decomposition (3.5), we obtain

$$\Delta X_t = \Pi(\nu_{t-1}) + \sum_{j=1}^{p-1} \tilde{\Phi}_j(\Delta X_{t-j}) + \Pi(X_0 - \nu_0) + \varepsilon_t, \quad t \geq p. \tag{4.4}$$

The sequence $(\Delta X_t - \Pi(\nu_{t-1}) - \sum_{j=1}^{p-1} \tilde{\Phi}_j(\Delta X_{t-j}) - \varepsilon_t, t \geq p)$ is a standard linear process with innovations ε. It therefore satisfies a law of large numbers (Bosq, 2000, p. 194) such that $n^{-1} \sum_{t=1}^{n}(\Delta X_t - \Pi(\nu_{t-1}) - \sum_{j=1}^{p-1} \tilde{\Phi}_j(\Delta X_{t-j}) - \varepsilon_t) \to 0$ almost surely. But from (4.4) we see that our summands are all equal to $\Pi(X_0 - \nu_0)$, which does not depend on t, and so we must have $\Pi(X_0 - \nu_0) = 0$. □

Lemma 4.3. *Under Assumption 4.1 we have* $\Psi_0 = \mathrm{id}_H$ *and, for all* $k \geq 1$,

$$\Psi_k = \Pi \tilde{\Psi}_{k-1} + \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi_{k-j}, \tag{4.5}$$

where we define $\Psi_k = 0$ *for* $k < 0$.

Proof. In view of Lemma 4.2 we have from (4.4) that

$$\Delta X_t = \Pi(\nu_{t-1}) + \sum_{j=1}^{p-1} \tilde{\Phi}_j(\Delta X_{t-j}) + \varepsilon_t, \quad t \geq p. \tag{4.6}$$

Writing the standard linear processes ΔX and ν in terms of their innovations ε, and rearranging terms, we deduce from (4.6) that

$$(\Psi_0 - \mathrm{id}_H)(\varepsilon_t) + \sum_{k=1}^{\infty} \left(\Psi_k - \Pi \tilde{\Psi}_{k-1} - \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi_{k-j} \right)(\varepsilon_{t-k}) = 0, \quad t \geq p. \tag{4.7}$$

Taking the covariance operator of both sides of (4.7) yields

$$
\begin{aligned}
0 &= (\Psi_0 - \mathrm{id}_H)\Sigma(\Psi_0 - \mathrm{id}_H)^* \\
&\quad + \sum_{k=1}^{\infty}\left(\Psi_k - \Pi\tilde{\Psi}_{k-1} - \sum_{j=1}^{p-1}\tilde{\Phi}_j\Psi_{k-j}\right)\Sigma\left(\Psi_k - \Pi\tilde{\Psi}_{k-1} - \sum_{j=1}^{p-1}\tilde{\Phi}_j\Psi_{k-j}\right)^* . \quad (4.8)
\end{aligned}
$$

Since Σ is positive definite, the desired result follows. □

Remark 4.5. Equation (4.5) is given by Hansen (2005, Theorem 1 and Lemma A.5) in the classical case $H = \mathbb{R}^n$.

Lemma 4.4. *Under Assumption 4.1 we have* $\Pi(\mathrm{id}_H + \Pi)^{k-1} = \Pi\tilde{\Psi}_{k-1}$ *for all* $k \geq 1$.

Proof. Recursive backward substitution in the error correction representation (4.2) yields, for any $t \geq p$,

$$
\begin{aligned}
\Delta X_t &= \varepsilon_t + \Pi(\varepsilon_{t-1}) + \sum_{j=1}^{p-1}\tilde{\Phi}_j(\Delta X_{t-j}) + \Pi(\mathrm{id}_H + \Pi)(X_{t-2}) \quad (4.9)\\
&= \varepsilon_t + \sum_{k=1}^{t-1}\Pi(\mathrm{id}_H + \Pi)^{k-1}(\varepsilon_{t-k}) + \sum_{j=1}^{p-1}\tilde{\Phi}_j(\Delta X_{t-j}) \\
&\quad + \Pi(\mathrm{id}_H + \Pi)^{t-1}(X_0). \quad (4.10)
\end{aligned}
$$

In view of Lemma 4.2 we thus obtain

$$
\begin{aligned}
\Delta X_t &= \varepsilon_t + \sum_{k=1}^{t-1}\Pi(\mathrm{id}_H + \Pi)^{k-1}(\varepsilon_{t-k}) + \sum_{j=1}^{p-1}\tilde{\Phi}_j(\Delta X_{t-j}) \\
&\quad + \Pi(\mathrm{id}_H + \Pi)^{t-1}(\nu_0) \quad (4.11)\\
&= \varepsilon_t + \sum_{k=1}^{t-1}\Pi(\mathrm{id}_H + \Pi)^{k-1}(\varepsilon_{t-k}) + \sum_{j=1}^{p-1}\tilde{\Phi}_j(\Delta X_{t-j}) \\
&\quad + \sum_{k=t}^{\infty}\Pi(\mathrm{id}_H + \Pi)^{t-1}\tilde{\Psi}_{k-t}(\varepsilon_{t-k}). \quad (4.12)
\end{aligned}
$$

Writing ΔX_t as a standard linear process with innovations ε and subtracting the right hand side of (4.12) from the left gives, for $t \geq p$,

$$
\begin{aligned}
0 \;=\; & (\Psi_0 - \mathrm{id}_H)(\varepsilon_t) + \sum_{k=1}^{t-1} \left(\Psi_k - \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi_{k-j} - \Pi(\mathrm{id}_H + \Pi)^{k-1} \right)(\varepsilon_{t-k}) \\
& + \sum_{k=t}^{\infty} \left(\Psi_k - \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi_{k-j} - \Pi(\mathrm{id}_H + \Pi)^{t-1} \tilde{\Psi}_{k-t} \right)(\varepsilon_{t-k}). \qquad (4.13)
\end{aligned}
$$

Since Σ is positive definite and t may be chosen arbitrarily large, examination of the covariance operator of the right-hand side of (4.13) reveals that

$$
\Pi(\mathrm{id}_H + \Pi)^{k-1} = \Psi_k - \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi_{k-j}, \quad k \geq 1. \qquad (4.14)
$$

Our desired result now follows from Lemma 4.3. □

Lemma 4.5. *Under Assumption 4.1 we have* $\ker \Pi \subseteq \cap_{k \geq 0} \ker \Pi \tilde{\Psi}_k$.

Proof. Immediate from Lemma 4.4. □

Proof of Theorem 4.1. The second assertion of our theorem follows from the first assertion and (2.7). In view of Lemma 4.1, to show the first assertion, it suffices to show that $\ker \Pi \subseteq (\ker \Lambda)^{\perp}$.

For each $x \in \ker \Pi$, Lemmas 4.3 and 4.5 jointly imply that

$$
\Psi_k(x) = \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi_{k-j}(x), \quad k \geq 1. \qquad (4.15)
$$

Summing both sides of (4.15) over $k \geq 1$ and using the fact that $\Psi_0 = \mathrm{id}_H$ (see Lemma 4.3) we obtain $\Psi(x) - x = \sum_{j=1}^{p-1} \tilde{\Phi}_j \Psi(x)$. It follows that

$$
\Gamma \Psi(x) = x, \quad x \in \ker \Pi, \qquad (4.16)
$$

where we have defined

$$
\Gamma = \mathrm{id}_H - \sum_{j=1}^{p-1} \tilde{\Phi}_j. \qquad (4.17)
$$

From (4.16) we see that the restriction of Ψ to $\ker \Pi$ is left invertible, and hence injective. The operator

$$
\bar{\Psi} : \ker \Pi \to \Psi(\ker \Pi), \qquad (4.18)
$$

defined by the requirement that $\bar{\Psi}(x) = \Psi(x)$ for all $x \in \ker \Pi$, is thus bijective.

We next show that the linear spaces $\ker \Pi$ and $\Psi(\ker \Pi)$ are equal to one another. Since $\bar{\Psi}$ is a linear bijection between the two spaces, they must have the same dimension. That dimension is finite by assumption, so to show that $\ker \Pi = \Psi(\ker \Pi)$ it suffices to show that $\Psi(\ker \Pi) \subseteq \ker \Pi$. Suppose $y \in \Psi(\ker \Pi)$. Then there exists $x \in \ker \Pi$ such that $y = \Psi(x)$, and thus since $\Psi = \mathrm{id}_H - \tilde{\Psi}_0$ we obtain

$$\Pi(y) = \Pi(x) - \Pi\tilde{\Psi}_0(x). \tag{4.19}$$

Since $x \in \ker \Pi$, the first term on the right-hand side of (4.19) is equal to zero, and in view of Lemma 4.5 the second term on the right-hand side of (4.19) is also equal to zero. Thus $y \in \ker \Pi$, which shows that $\Psi(\ker \Pi) \subseteq \ker \Pi$.

We have shown that $\bar{\Psi}$ is an invertible map from $\ker \Pi$ to $\ker \Pi$. It follows that the adjoint operator $\bar{\Psi}^*$, which satisfies $\bar{\Psi}^*(y) = \Psi^*(y)$ for all $y \in \ker \Pi$, is also an invertible map from $\ker \Pi$ to $\ker \Pi$. Thus, $\ker \bar{\Psi}^* = \{0\}$, implying that $(\ker \Psi^*) \cap (\ker \Pi) = \{0\}$, and consequently in view of (3.3) we have

$$(\ker \Lambda) \cap (\ker \Pi) = \{0\}. \tag{4.20}$$

The intersection of the orthogonal complements of two linear spaces is equal to the orthogonal complement of their sum (Conway, 1990, p. 40, Ex. 3), and so from (4.20) we have $((\ker \Lambda)^\perp + (\ker \Pi)^\perp)^\perp = \{0\}$. Taking the orthogonal complement of both sides gives

$$\mathrm{cl}((\ker \Lambda)^\perp + (\ker \Pi)^\perp) = H. \tag{4.21}$$

Fix an arbitrary $z \in \ker \Lambda$. Equation (4.21) implies the existence of a sequence $(x_n, n \geq 1)$ in $(\ker \Lambda)^\perp$ and a sequence $(y_n, n \geq 1)$ in $(\ker \Pi)^\perp$ such that $x_n + y_n \to z$ as $n \to \infty$. Observe that

$$\|(x_n + y_n) - z\|^2 = \|x_n\|^2 + \|y_n - z\|^2 + 2\langle x_n, y_n - z\rangle. \tag{4.22}$$

Lemma 4.1 implies that $y_n \in \ker \Lambda$ for each $n \geq 1$. Therefore, the inner product on the right-hand side of (4.22) is equal to zero for each $n \geq 1$. Since the term on the left-hand side of (4.22) is converging to zero as $n \to \infty$, it follows that we must have $\|y_n - z\| \to 0$ as $n \to \infty$. But $z \in \ker \Lambda$ was arbitrary, so we must have $\ker \Lambda \subseteq \mathrm{cl}(\ker \Pi)^\perp = (\ker \Pi)^\perp$. It follows that $\ker \Pi \subseteq (\ker \Lambda)^\perp$, as claimed. □

References

Beveridge, S., and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle.' *Journal of Monetary Economics*, **7**, 151–174.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Springer, New York.

Chang, Y., Kim, C. S. and Park, J. Y. (2015). Nonstationarity in time series of densities. *Journal of Econometrics*, in press.

Conway, J. B. (1990). *A Course in Functional Analysis*, 2nd ed. Springer, New York.

Davidson, J. (2009). When is a time series I(0)? Ch. 13 in Castle, J. and Shepherd, N. (Eds.), *The Methodology and Practice of Econometrics: A Festschrift for David F. Hendry*, pp. 322–342. Oxford University Press, Oxford.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica*, **55**, 251–276.

Gabrys, R., Hörmann, S. and Kokoszka, P. (2013). Monitoring the intraday volatility pattern. *Journal of Time Series Econometrics*, **5**, 87–116.

Granger, C. W. J. (1981). Some properties of times series data and their use in econometric model specification. *Journal of Econometrics*, **16**, 121–130.

Granger, C. W. J. (1983). Cointegrated variables and error-correcting models. Mimeo, Department of Economics, UC San Diego.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton.

Hansen, P. R. (2005). Granger's representation theorem: A closed-form expression for $I(1)$ processes. *Econometrics Journal*, **8**, 23–28.

Hörmann, S. and Kokoszka, P. (2012). Functional time series. Ch. 7 in Rao, T. S., Rao, S. S. and Rao, C. R. (Eds.), *Handbook of Statistics, Vol. 30: Time Series Analysis—Methods and Applications*, pp. 157–186. North-Holland, Amsterdam.

Hörmann, S., Horváth, L. and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory*, **29**, 267–288.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications.* Springer, New York.

Johansen, S. (1988). The mathematical structure of error correction models. *Contemporary Mathematics*, **80**, 359–386.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, **59**, 1551–1580.

Johansen, S. (1992). A representation of vector autoregressive processes integrated of order 2. *Econometric Theory*, **8**, 188–202.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford University Press, Oxford.

Johansen, S. (2009). Representation of cointegrated autoregressive processes with application to fractional processes. *Econometric Reviews*, **28**, 121–145.

Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, **99**, 2508–2526.

Kokoszka, P. and Zhang, X. (2012). Functional prediction of intraday cumulative returns. *Statistical Modelling*, **12**, 377–398.

Park, J. Y. and Qian, J. (2012). Functional regression of continuous state distributions. *Journal of Econometrics*, **167**, 397–412.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, **20**, 971–1001.

Stock, J. H. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association*, **83**, 1097–1107.