Time Series
Steven N. Durlauf
University of Wisconsin

Lecture Notes 2. The Hilbert Space Approach to Time Series

1. Basic ideas

The Hilbert space framework provides a very powerful language for discussing the relationship between various random variables. Collections of random variables are called *stochastic processes*; in common usage stochastic processes are usually understood to refer to collections of random variables whose elements are indexed by time. We focus on the case of a scalar stochastic process $\{x_t\}$ where *t* is an integer as it is convenient to think of time stretching from $-\infty$ to ∞. We assume that the process is zero-mean and second-order stationary. Second-order stationarity, also known as weak stationarity, means that the autocovariances between x_{t-j} and x_t do not depend on *t.* Formally, assume i) $E(x_t)=0$ and ii) $E(x_t x_{t-j})=\sigma(j)<\infty$.

For random variables such as the elements of the stochastic process x_t, the natural metric, i.e. notion of length, for a random variable its standard deviation,

$$\|x_t\|=\sqrt{\mathrm{E}\left(x_t^2\right)} \tag{2.1}$$

with covariance as the associated notion of inner product,

$$\langle x_t, x_{t-j}\rangle = \mathrm{E}\left(x_t x_{t-j}\right) \tag{2.2}$$

One can generate a Hilbert space around the sequence. $\{x_t, x_{t-1}, x_{t-2}\ldots\}$. What this means is that one forms a space by taking these elements, adding all linear combinations of the elements, all limits of the linear combinations, etc[1]. We denote this Hilbert space as $H_t(x)$. The entire history of the stochastic process from $-\infty$ to ∞ generates $H_\infty(x)$. By construction, $H_{t-1}(x) \subseteq H_t(x)$.

The general properties of Hilbert spaces described in Lecture Notes 1 allow one to characterize the linear structure of $H_t(x)$ in ways that are very important in macroeconomics. First, observe that by the Hilbert space decomposition theorem, one can decompose $H_t(x)$ so that

$$H_t(x) = H_{t-1}(x) \oplus G_t \tag{2.3}$$

where G_t is another Hilbert space. The dimension of this Hilbert space is either 0 or 1. This is so because the Hilbert space G_t must be spanned by the single random variable $x_t - proj\left(x_t \middle| H_{t-1}(x)\right)$ where $proj\left(x_t \middle| H_{t-1}(x)\right)$ is the projection of x_t onto $H_{t-1}(x)$. To say the space G_t has dimension 0 means that $x_t \in H_{t-1}(x)$, i.e. $\text{var}\left(x_t - proj\left(x_t \middle| H_{t-1}(x)\right)\right) = 0$. If one again applies the Hilbert space decomposition theorem, it is the case that

$$H_t(x) = H_{t-2}(x) \oplus G_{t-1} \oplus G_t \tag{2.4}$$

Here G_{t-1} is spanned by $x_{t-1} - proj\left(x_{t-1} \middle| H_{t-2}(x)\right)$. One can repeat this decomposition any number of times. The G_t spaces are by construction mutually orthogonal. This may be repeated an arbitrary number of times.

Notice that it is not the necessarily the case that, if this decomposition is done an infinite number of times, the G_t's may be used to reconstruct $H_t(x)$. The reason for this is each space is constituted by elements that appear in the

---

[1]Technically, we are working with the smallest Hilbert space that contains the

space $H_{t-j}(x)$ but not in the space $H_{t-j-1}(x)$; if there are elements that appear in every member of the sequence $H_t(x), H_{t-1}(x)$…, they will not appear in any of the G_t's. Elements that are common to all of the $H_t(x)$'s form a Hilbert space as well. Formally, this space is defined as

$$H_{-\infty}(x) = \bigcap_{t=-\infty}^{\infty} H_t(x) \tag{2.5}$$

The Hilbert space generated by current and past x_t's can therefore be decomposed as

$$H_t(x) = G_t \oplus G_{t-1} \oplus \ldots \oplus H_{-\infty}(x) \tag{2.6}$$

2. Wold decomposition theorems

Equation (2.6) is the basis for two fundamental theorems in time series analysis, each due to Herman Wold; his 1948 article is still worth reading. Rozanov (1967) is a deep treatment. I find Ash and Gardner's (1967) discussion to be especially insightful.

Theorem 2.1. Wold decomposition theorem I

Any zero-mean, finite variance, second-order stationary x_t may be decomposed as

$$x_t = x_{1t} + x_{2t} \tag{2.7}$$

---

elements of the stochastic process.

where

$$x_{1t} \in G_t \oplus G_{t-1} \oplus G_{t-2} \oplus \ldots \tag{2.8}$$

and

$$x_{2t} \in H_{-\infty}(x) \tag{2.9}$$

In this decomposition, x_{1t} is called the indeterministic component and x_{2t} the deterministic component of x_t. The terms refer to whether or not the components may be perfectly (linearly) predicted from the past. When a time series contains a nontrivial indeterministic component, the time series itself is said to be indeterministic. If the process does not contain a deterministic component, it is purely indeterministic. The term x_{2t} may be perfectly predicted from information in the arbitrarily distant past.

One example of deterministic component is a seasonal. Consider the stochastic process $\cos(\omega t + \theta)$ where θ is uniformly distributed on $[-\pi, \pi]$. Since $\int_{-\pi}^{\pi} \cos(\omega t + \theta) d\theta = 0$ and $\int_{-\pi}^{\pi} \cos(\omega(t-j) + \theta)\cos(\omega t + \theta) d\theta$ does not depend on t,[2] $\cos(\omega t + \theta)$ is a candidate for x_{2t}.

The second Wold Theorem characterizes the linear structure of the indeterministic part of a time series.

Theorem 2.2. Wold decomposition theorem II

[2]This follows from the identity

$$\cos(\omega(t-j) + \theta)\cos(\omega t + \theta) = \frac{1}{2}\left(\cos(j\omega) + \cos(2\omega t + j\omega + 2\theta)\right).$$

If x_t is a purely indeterministic, zero-mean, finite variance, second-order stationary process, then there exists a representation of the process of the form

$$x_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} \ , \ \alpha_0 = 1 \tag{2.10}$$

where $\varepsilon_t \in G_t$ and $\sigma^2_{\varepsilon_t} = \sigma^2_{\varepsilon_{t-j}} \ \forall j$. $\sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ is known as the fundamental moving average (MA) representation of x_t and is unique.

Pf. Since $H_t(x) = H_{t-1}(x) \oplus G_t$, by construction G_t is a Hilbert space of maximum dimension 1; the space is spanned by ε_t which is a scalar random variable. If the dimension is zero (which would mean that $\text{var}(\varepsilon_t) = 0$), then $H_t(x) = H_{t-1}(x)$ which contradicts the assumption that the process is purely indeterministic. Since the process is indeterministic, one may find an element ε_t in G_t such that the projection of x_t onto G_t is ε_t; this is nothing more than a choice of axis for the one dimensional space. For the spaces G_{t-j} ($j > 0$), one can find an element in each of them, denoted as ε_{t-j}, whose variance equals that of ε_t; each ε_{t-j} spans its respective space. Since x_t is purely indeterministic, $H_t(x) = G_t \oplus G_{t-1} \oplus ...$ Letting $proj\left(x_t \middle| G_{t-j}\right)$ denote the projection of x_t onto G_{t-j}, by the Hilbert space projection theorem

$$x_t = \sum_{j=0}^{\infty} proj\left(x_t \middle| G_{t-j}\right) = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j} \tag{2.11}$$

where the second equality follows from the definition of the ε_t's. This verifies the theorem, except for uniqueness.

To prove uniqueness, suppose that there existed another MA representation $x_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j.}$ For this to be the case, the variance of $\sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j} - \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$ must equal zero since by assumption the parts of the expression are the same. The variance of this expression equals $\sigma_\varepsilon^2 \sum_{j=0}^{\infty} \left(\alpha_j - \beta_j\right)^2$, which equals zero iff $\alpha_j - \beta_j = 0 \; \forall \; j$.

An immediate implication of the second Wold theorem is that the fundamental moving average coefficients are square summable, i.e. $\sum_{j=0}^{\infty} \alpha_j^2 < \infty$. This follows from the fact that $\text{var}\left(x_t\right) = \text{var} \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t=j} = \sigma_\varepsilon^2 \sum_{j-0}^{\infty} \alpha_j^2$ and the assumption that $\sigma_x\left(0\right) < \infty$.

What is meant by the term fundamental in the description of the moving average representation? It is a way of designating a particular orthogonal basis for $H_t\left(x\right)$. There exist an uncountable infinity of different orthogonal bases for $H_t\left(x\right)$, just as there are for spaces such as R^k. As far as I know the term is taken from Rozanov (1967); it was popularized by Christopher Sims.

There is an equivalence between the Hilbert space generated around the stochastic process x_t and the Hilbert space generated around the fundamental innovations ε_t

Theorem 2.3. Equivalence between the Hilbert space of a time series and its associated fundamental innovations.

Let $H_t\left(\varepsilon\right)$ denote the Hilbert space generated by $\varepsilon_t, \varepsilon_{t-1} \ldots$, the fundamental moving average components of a zero-mean, weakly stationary process x_t. Then $H_t\left(\varepsilon\right) = H_t\left(x\right)$.

Pf. This is left as an exercise; the proof amounts to showing that each of the two Hilbert spaces is a subset of the other.

3. Prediction

Finally, we consider the question of how to optimally predict a time series given its history. Let $x_{t|t-j}$ denote the projection of x_t onto $H_{t-j}(x)$. This projection is important in that it is also the solution to the linear prediction problem for x_t relative to the information set $H_{t-j}(x)$.[3]

Theorem 2.4. Optimal linear predictor.

The projection $x_{t|t-j}$ is the solution to $\min_{\xi \in H_{t-j}(x)} E\left(x_t - \xi\right)^2$

Pf. Let $\bar{\xi}$ solve the minimization problem. The prediction error equals $\varepsilon_t + x_{t|t-j} - \bar{\xi}$. The variance of this term will equal $\sigma_\varepsilon^2 + \sigma_{x_{t|t-j}-\bar{\xi}}^2$, since ε_t is orthogonal to $x_{t|t-j} - \bar{\xi}$. This variance must exceed σ_ε^2 unless $x_{t|t-j} - \bar{\xi}$ is zero. Uniqueness of the projection then verifies the result.

This theorem implies that ε_t is the forecast error associated with the optimal (in a minimum variance sense) linear forecast of x_t given the information set $H_{t-1}(x)$. The term linear means that the forecast has to be an element of the space, and so is either a linear combination of x_{t-1}, x_{t-2}, … or the limit of such a

[3]This theorem a special case of the general result on the relationship between Hilbert space projections and certain minimization problems described in Lecture Notes 1.

sequence. Hence, one can think of a time series as a weighted average of current and past forecast errors. This is intuitive since these forecast errors reveal aspects of the process that are realized each time period.

The optimal linear predictor theorem provides insight into the nature of information sets and associated predictions. By construction, $x_{t|t-j} \in H_{t-j}(x)$ and $x_t - x_{t|t-j} \in G_t \oplus G_{t-1} \oplus ... \oplus G_{t-j+1}$. The fundamental innovations $\varepsilon_t, ..., \varepsilon_{t-j+1}$ represent the part of x_t that is revealed after the forecast. The fundamental innovations can be thought of as information increments between time periods.

From the perspective of the Wold theorems, the moving average representation of a time series can thus be seen as a natural way of thinking about its underlying linear structure; much of time series analysis is based on this perspective, which is often referred to as the time domain approach. That said, nothing in these derivations rules out the presence of nonlinear structure in the stochastic process.

References

Ash, R. and M. Gardner, (1975), *Topics in Stochastic Processes.* New York: Academic Press.

Rozanov, Y., (1967), *Stationary Time Series*, San Francisco: Holden Day.

Wold, H., (1948), "On Prediction in Stationary Time Series," *Annals of Mathematical Statistics*, 19, 4, 558-567.