FUNCTIONAL GRANGER CAUSATION

MARCELO M. TADDEO[1] AND RAYDONAL OSPINA[1]

ABSTRACT. ....

Keywords. Recurrent events, splines, empirical processes.

1. INTRODUCTION

....

2. PRELIMINARIES

2.1. Banach and Hilbert Space Valued Random Elements.

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a given complete probability space and $\mathbb{B}$ a separable Banach space with norm $\|\cdot\|_{\mathbb{B}}$ and equipped with the Borel σ-field $\mathcal{B}(\mathbb{B})$. Since we are primarily concerned with $\mathbb{B}$-valued stochastic processes, we summarize the main concepts we will need.

Definition 1. *A map $X : \Omega \to \mathbb{B}$ is said to be a weakly measurable function if $\psi(X) : \Omega \to \mathbb{R}$ is measurable for every $\psi \in \mathbb{B}'$, where $\mathbb{B}'$ is the topological dual of $\mathbb{B}$. Such functions will be referred to as $\mathbb{B}$-random elements.*

The assumption that X is weakly measurable allows us to consider the following (weak) integral for $\mathbb{B}$-valued functions.

Definition 2. *Any weakly measurable function $X : \Omega \to \mathbb{B}$ satisfying (i) $\psi(X) \in L^1(\Omega, \mathcal{F}, \mathrm{P})$, for all $\psi \in \mathbb{B}'$, and (ii) there is $Y \in \mathbb{B}$ such that, for all $\psi \in \mathbb{B}'$, $\psi(Y) = \int \psi(X) d\,\mathrm{P}$, is said to be weakly integrable. The random element Y is denoted by $\int X d\,\mathrm{P}$ and referred to as the weak integral of X.*

Date: September 6, 2023.

Additionally, if $\mathrm{E}\,\|X\|_{\mathbb{B}} < \infty$, we say that X is integrable. In particular, if X is integrable, then X is also strongly (Bochner) integrable. Actually, if X is integrable, then the notions of weakly and strongly integrable functions are equivalent. This follows from the fact that every integrable X can be approximated by a sequence of simple functions. The reason we use weak integrals instead is that they are more suitable to deal with the second moments of $\mathbb{B}$-random elements. Since P is a probability measure it is natural to identify the weak integral of X with its (weak) expectation, which will be denoted as usual by $\mathrm{E}\,X$. The expectation $\mathrm{E}\,X$ is associated with the operator $\mathrm{E}_X\,\psi = \int \psi(X) d\,\mathrm{P}$, for $\psi \in \mathbb{B}'$, so that $\mathrm{E}\,X$ is such that $\psi(\mathrm{E}\,X) = \mathrm{E}_X\,\psi$, for all $\psi \in \mathbb{B}'$.

Proposition 1. *For any $X \in \mathbb{B}$, the operator $\mathrm{E}_X : \mathbb{B}' \to \mathbb{R}$ is linear and $\|\,\mathrm{E}_X\,\| = \mathrm{E}\,\|X\|$.*

Proof. Linearity follows trivially from the definition of E_X. Now, if $\psi \in \mathbb{B}'$,

$$|\,\mathrm{E}_X\,\psi| = |\mathrm{E}\,\psi(X)| \leq \mathrm{E}\,|\psi(X)| \leq \mathrm{E}\,\|\psi\|\|X\| = (\mathrm{E}\,\|\psi\|)\,\|X\|,$$

so that $\|\,\mathrm{E}_X\,\| \leq \mathrm{E}\,\|\psi\|$. On the other hand, let $\mathbb{B}_X = \langle X \rangle$, the subspace generated by $X \in \mathbb{B}$, and consider the functional (defined on $\mathbb{B}_X$) $\tilde{\psi}(\alpha X) = \alpha\|X\|$, which clearly is bounded linear with norm $\|\tilde{\psi}\| = 1$. Using the Hahn-Banach Theorem, $\tilde{\psi}$ can be extended to a bounded linear functional ψ on $\mathbb{B}$ such that $\|\psi\| = \|\tilde{\psi}\| = 1$. In particular, $\psi(X) = \tilde{\psi}(X) = \|X\|$. Hence,

$$\|\,\mathrm{E}_X\,\| \geq |\,\mathrm{E}_X\,\psi| = \left| \int \psi(X) d\,\mathrm{P} \right| = \left| \int \|X\| d\,\mathrm{P} \right| = \mathrm{E}\,\|X\|,$$

so that $\|\,\mathrm{E}_X\,\| = \mathrm{E}\,\|X\|$. □

In other words, $\mathrm{E}_X \in \mathbb{B}''$, the topological second dual of $\mathbb{B}$. A similar argument as in the proof of Proposition 1 shows us that $\|\,\mathrm{E}_X\,\| = \|\,\mathrm{E}\,X\|$. For Hilbert spaces $\mathbb{H}$, it follows from Riesz's theorem that there exists an unique $Y \in \mathbb{H}$ such that $\psi(X) = \langle X, Y \rangle$, for every $X \in \mathbb{H}$. Hence, $\mathrm{E}\,X$ is the element in $\mathbb{H}$ such that $\langle \mathrm{E}\,X, Y \rangle = \int \langle X, Y \rangle d\,\mathrm{P}$, for all $Y \in \mathbb{H}$. Moreover, $\mathrm{E}\,X$ is also the representer of E_X, since $\mathrm{E}_X\,\psi = \langle \mathrm{E}\,X, Y \rangle$, where $Y \in \mathbb{H}$ is the representer of ψ. Finally, notice that $\int[\psi(X - \mathrm{E}\,X)]d\,\mathrm{P} = \int \psi(X) - \psi(\mathrm{E}\,X) = 0$, for all $\psi \in \mathbb{B}'$, so that $\mathrm{E}[X - \mathrm{E}\,X] = 0$. Hence, without loss of generality, we may assume $\mathrm{E}\,X = 0$.

To define an analogous of a covariance measure for Banach space valued random elements, we consider a bilinear version of E_Φ. Given any pair of random elements $X_1 \in \mathbb{B}_1$, $X_2 \in \mathbb{B}_2$, with $\mathrm{E}\,X_1 = 0$, $\mathrm{E}\,X_2 = 0$ and $\psi(X_i) \in L^2(\Omega, \mathcal{F}, \mathrm{P})$, for any $\psi \in \mathbb{B}_i'$, define the covariance functional $\mathrm{C}_{X_1,X_2}(\psi_1, \psi_2) = \int \psi_1(X_1)\psi_2(X_2)d\,\mathrm{P} = \mathrm{Cov}(\psi_1(X_1), \psi_2(X_2))$ on $\mathbb{B}_1' \times \mathbb{B}_2'$. The functional C_{X_1,X_2} is clearly linear with $|\,\mathrm{C}_{X_1,X_2}(\psi_1, \psi_2)| \leq \mathrm{E}[\|X_1\|\|X_2\|]\|\psi_1\|\|\psi_2\|$ and it is connected to E_X in the following sense: $\mathrm{C}_{X_1,X_2}(\psi_1, \psi_2) = \int \psi_1(\psi_2(X_2)X_1)d\,\mathrm{P} = \mathrm{E}_{\psi_2(X_2)X_1}\,\psi_1 = \mathrm{E}_{\psi_1(X_1)X_2}\,\psi_2$. Conversely, given $\psi_2 \in \mathbb{B}_2'$, $\mathrm{E}_{\psi_1(X_1)X_2}\,\psi_2 = \psi_2(\mathrm{E}[\psi_1(X_1)X_2])$, meaning that $\mathrm{C}_{X_1,X_2}(\psi_1, \psi_2)$ is completely determined by the cross-covariance operators $c_{X_1,X_2} : \psi \in \mathbb{B}_1' \mapsto \mathrm{E}[\psi(X_1)X_2] \in \mathbb{B}_2$. If $X_1 = X_2 = X \in \mathbb{B}$, we write $\mathrm{C}_{X_1,X_2}(\psi_1, \psi_2) = \mathrm{C}_X(\psi_1, \psi_2)$. In this case, the covariance functional is also symmetric positive semidefinite and, by the Schwarz inequality, $|\,\mathrm{C}_X(\psi_1, \psi_2)|^2 \leq \mathrm{V}_X(\psi_1)\,\mathrm{V}_X(\psi_2)$, where $\mathrm{V}_X(\psi) \equiv \mathrm{C}_X(\psi, \psi)$, called the variance functional for X. Both $C_X(\psi, \psi)$ and $V_X(\psi)$ are completely determined by $c_X(\psi) \equiv c_{X,X}(\psi) = \mathrm{E}[\psi(X)X]$ in $\mathbb{B}$. For Hilbert spaces, all these functionals can be clearly written in terms of inner products.

2.2. **Functional FARMA.** A sequence $\varepsilon = (\varepsilon_t : t \in \mathbb{Z})$ of zero mean $\mathbb{B}$-random elements is said to be a strong white noise if (i) its elements are i.i.d., (ii) $0 < \mathrm{E}\,\|\varepsilon_t\|^2 = \sigma^2 < \infty$ and (iii) $\mathrm{E}\,\varepsilon_t = 0$. A sequence $X = (X_t : t \in \mathbb{Z})$ of $\mathbb{B}$-random elements is a strong functional autorregressive model of order p if

$$\sum_{j=0}^{p} \Phi_j(X_{t-j} - \mu) = \varepsilon_t,$$

where $\mu \in \mathbb{B}$, ε is a strong white noise and $\Phi_0, ..., \Phi_p \in \mathcal{L}(\mathbb{B}, \mathbb{B})$, with $\Phi_0 = I$ and $\Phi_p \neq 0$. If we relax the condition (i) of independence between the elements ε_t for (i') the covariance operators c_{ε_t} do not depend on t and $c_{\varepsilon_t,\varepsilon_s} \equiv 0$ if $t \neq s$, we obtain what is called weak white noise. The sequence X above is referred to as a weak autorregressive process. Whenever this distinction is unnecessary, X will

Se $p = 1$, a condicao (iii) não eh imposta — **?**, p. 127.

be referred to just as a autorregressive process. In any case we will write $X \sim \mathrm{FAR}(p)$. The extension to functional ARMA (FARMA) models is straighforward. In fact, a sequence $X = (X_t : t \in \mathbb{Z})$ of $\mathbb{B}$-random elements is said to be a (weak) strong functional ARMA process in $\mathbb{B}$ of order p and q if

$$\sum_{j=0}^{p} \Phi_j(X_{t-j} - \mu) = \sum_{j=0}^{q} \Psi_j \varepsilon_{t-j},$$

where μ, Φ_j are as before, ε is a (weak) strong white noise and $\Psi_0, ..., \Psi_q \in \mathcal{L}(\mathbb{B}, \mathbb{B})$, with $\Psi_0 = I$ and $\Psi_q \neq 0$. In this case, we use the notation $X \sim \mathrm{FARMA}(p, q)$.

2.3. **Linear Prediction.** A subset $\mathcal{L}_o$ of $L_{\mathbb{B}}(\Omega, \mathcal{F}, \mathrm{P})$ is a $\mathcal{L}$-closed subspace if it is a closed subspace and invariant under any linear operator in $\mathcal{L}$. Given a $\mathbb{B}$-valued zero-mean stationary process, let $\mathcal{M}_n^X$ be the $\mathcal{L}$-closed subspace generated by $\{X_k : k \leq n\}$, which means that $\mathcal{M}_n^X$ is just the closure of $\{\sum_{j=0}^{M} \Phi_j X_{n-j} : \Phi_j \in \mathcal{L},\ M \in \mathbb{N}\}$. The projection of X_{n+k} over $\mathcal{M}_n^X$, denoted by $\pi(X_{n+k}|\mathcal{M}_n^X)$, will be referred to as best linear k-step predictor of X_n and the corresponding k-step predictive error is defined by $\mathcal{E}_k(X|\mathcal{M}_n^X) = X_{t+k} - \pi(X_{n+k}|\mathcal{M}_n^X)$.

3. Multiple FAR(1) Processes

Given the Banach spaces $(\mathbb{B}_1, \|\cdot\|_1)$ and $(\mathbb{B}_2, \|\cdot\|_2)$, let $\mathbb{B} = \mathbb{B}_1 \oplus \mathbb{B}_2$ be their (external) direct sum equipped with the norm $\|(X_1, X_2)\| = (\|X_1\|_1^2 + \|X_2\|_2^2)^{1/2}$. In this section we are interested in Banach space valued processes of the form

$$\boldsymbol{X}_{t+1} - \boldsymbol{\mu} = \Phi(\boldsymbol{X}_t - \boldsymbol{\mu}) + \boldsymbol{\varepsilon}_{t+1}, \tag{1}$$

where $\boldsymbol{X}_t = (X_{1,t}, X_{2,t})^\top$, $\boldsymbol{\mu} = (\mu_1, \mu_2)^\top \in \mathbb{B}$, $\Phi : \mathbb{B} \to \mathbb{B}$ is a bounded linear operator, $X_{i,t}$ and $\varepsilon_{i,t}$ are $\mathbb{B}_i$-random elements, for $i = 1, 2$ and $t \in \mathbb{Z}$, and the sequence $(\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})^\top : t \in \mathbb{Z})$ is a strong white noise. Since $\mathrm{E}\,\|(\varepsilon_{1,t}, \varepsilon_{2,t})\|^2 = \mathrm{E}\,\|\varepsilon_{1,t}\|_{\mathbb{B}_1}^2 + \mathrm{E}\,\|\varepsilon_{2,t}\|_{\mathbb{B}_2}^2$, we may write $\sigma^2 = \sigma_1^2 + \sigma_2^2$, where $\sigma_1^2 \equiv \mathrm{E}\,\|\varepsilon_{1,t}\|_{\mathbb{B}_1}^2 < \infty$ and $\sigma_2^2 \equiv \mathrm{E}\,\|\varepsilon_{2,t}\|_{\mathbb{B}_2}^2 < \infty$. Moreover, each sequence $(\varepsilon_{i,t} : t \in \mathbb{Z})$ is itself a strong white noise. If the process $\boldsymbol{X}_t$ is stationary, it will be referred to as a multiple autorregressive process in $\mathbb{B}$ of order 1 and denoted by $\mathbb{B} - MAR(1)$. Of course, such a definition can be easily extended to more general processes by considering the space $\mathbb{B} = \mathbb{B}_1 \oplus \cdots \oplus \mathbb{B}_m$, but we will focus here on the case $m = 2$ just to keep notation simpler. Finally, we allow the elemens $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ to be correlated, but their covariance functionals must satisfy $C_{\varepsilon_{1,t},\varepsilon_{2,t}}(\psi_1, \psi_2) = C_{\varepsilon_{1,0},\varepsilon_{2,0}}(\psi_1, \psi_2)$, for all $t \in \mathbb{Z}$, $\psi_1 \in \mathbb{B}_1$ and $\psi_2 \in \mathbb{B}_2$. This is equivalent to assume $c_{\varepsilon_{i,t},\varepsilon_{j,t}} \equiv c_{\varepsilon_{i,0},\varepsilon_{j,0}}$, for any $t \in \mathbb{Z}$ and $i \neq j \in \{1, 2\}$.

We note now that the operator Φ can be written as

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{bmatrix}$$

where $\Phi_{ij} : \mathbb{B}_j \to \mathbb{B}_i$ is a bounded linear operator[(i)], for any $i, j = 1, 2$. Hence the $\mathbb{B} - MAR(1)$ process $\boldsymbol{X}_t$ can be written in the more usual form

$$X_{1,t+1} = \Phi_{11}X_{1,t} + \Phi_{12}X_{2,t} + \varepsilon_{1,t+1},$$

$$X_{2,t+1} = \Phi_{21}X_{1,t} + \Phi_{22}X_{2,t} + \varepsilon_{2,t+1}.$$

A ideia é avaliar quais condições devem satisfazer os elementos de Φ para que X_2 não Granger cause X_1.

Moreover, from Theorem 3.1 in **?**, if there exists $j_o \in \mathbb{N}$ such that $\|\Phi\|_{\mathcal{L}(\mathbb{B},\mathbb{B})} < 1$, then the process (1) can be uniquely represented by

$$\boldsymbol{X}_t = \boldsymbol{\mu} + \sum_{j \geq 0} \Phi^j \boldsymbol{\varepsilon}_{t-j},$$

for all $t \in \mathbb{Z}$.

By simple recursion, we get

$$(X_{1,t+1}, X_{2,t+1}) = \sum_{j=1}^{\infty} \Phi^j (\varepsilon_{1,t+1-j}, \varepsilon_{2,t+1-j}).$$

4. Functional Granger Causality

Loosely speaking, given two time series X and Y, it is said that X causes Y (in the Granger's sense) if X contains in itself relevant and unique information regarding Y. More precisely, if $\mathcal{J}_t$ stands for all the available information in universe at

(i) Indeed, since Φ is a bounded linear operator from $\mathbb{B}_1 \oplus \mathbb{B}_2$ to itself, then $\Phi(X_{1,t}, X_{2,t}) = (\Phi_1(X_{1,t}, X_{2,t}), \Phi_2(X_{1,t}, X_{2,t}))$, where Φ_i is a bounded linear operator from $\mathbb{B}_1 \oplus \mathbb{B}_2$ to $\mathbb{B}_i$. Now, if $\Phi_{i1}(X_{1,t}) = \Phi_i(X_{1,t}, 0)$ and $\Phi_{i2}(X_{2,t}) = \Phi_i(0, X_{2,t})$, then both are bounded linear operators and $\Phi_i(X_{1,t}, X_{2,t}) = \Phi_i(X_{1,t}, 0) + \Phi_i(0, X_{2,t}) = \Phi_{i1}(X_{1,t}) + \Phi_{i2}(X_{2,t})$.

t and $\mathcal{J}_t'$ stands for all the available information at t except for that contained in $\{Y_{t-k} : k \geq 0\}$, then X is said not to Granger-cause Y if

$$Y_{t+1} \perp\!\!\!\perp \mathcal{J}_t | \mathcal{J}_t',$$

for all t. Otherwise, X is said to Granger-cause Y. This definition has three problems. The first one concerning the idea of causality itself and the second related to a measure theoretic isse regarding the "size" of $\mathcal{J}_t'$ in the sense that for sufficiently rich processes X and Y, it is not true that $\mathcal{J}_t'$ is smaller that $\mathcal{J}_t$, even if X has actually caused Y in a more strict sense, see Eichler (2012). Finally, it is not practical to consider $\mathcal{J}_t$ to be all the available information in the universe and the independence condition (**??**). In practice, it is more convenient to consider $\mathcal{J}_t$ and $\mathcal{J}_t'$ to be the information generated by $\{(X_{t-k}, Y_{t-k}) : k \geq 1\}$ and $\{Y_{t-k} : k \geq 1\}$, respectively. Moreover, instead of independence, it is said that X does not Granger-cause Y if the predictions of Y, the first one given $\mathcal{J}$ and the second one given $\mathcal{J}'$, both result in the same prediction error.

This definition extends naturally for functional time series. In fact,

REFERENCES

Bosq, D. (2000). *Linear Processes in Function Spaces – Theory and Applications*. New York: Springer-Verlag.

Eichler, M. (2012). Causal inference in time series analysis. In Berzuini, C., Dawid, A.P., and Bernardinelli, L. (eds.), *Causality*, 327 – 354. John Wiley and Sons.

[1]DEPARTMENT OF STATISTICS, UNIVERSITY OF BAHIA, BRAZIL