Forecasting densities of financial returns in probability spaces

PPGE/UFRGS - Economia Aplicada

Requisito parcial para o título de Mestre em Economia

Matheus Vizzotto dos Santos - 00290240

12 de abril de 2025

1 Abstract

Functional time series -> density functions.

Hall & Vial -> Bathia et al.

Espaço de funções de densidade (integração, não negativa).

Dados utilizados.

2 Introdução

2.1 Density estimation

Comentar sobre estimação de funções de densidade de probabilidade.

Abordagens iniciais e mais recentes (machine learning etc.).

Pisar no assunto de dados de alta frequência (-> análise funcional).

2.2 A glimpse into Functional Data Analysis

Progressão de análise multivariada para análise funcional (exemplo PCA -> FPCA).

Trabalhos iniciais (ex Ramsay).

Gertheiss et al. (2024), Dabo-Niang e Frévent (2024)

2.3 Functional Time Series

Hall e Vial (2006), Bathia, Yao e Ziegelmann (2010), Aue, Norinho e Hörmann (2015), Bosq (2000)

Functional data analysis (FDA) is a branch of statistics that deals with data providing information about curves, surfaces or anything else varying over a continuum. In this context, a *functional time series* (FTS) is a sequence of random functions indexed by time. Each observation in the series is a function, typically lying in an infinite-dimensional function space, such as $L^2(\mathcal{T})$ for some compact interval $\mathcal{T} \subset \mathbb{R}$.

Let $\mathcal{H}$ be a separable Hilbert space, typically taken to be $\mathcal{H} = L^2(\mathcal{T})$, the space of square-integrable functions on a compact interval $\mathcal{T} \subset \mathbb{R}$, equipped with the inner product

$$\langle f, g \rangle = \int_{\mathcal{T}} f(t) g(t) \, dt,$$

and the associated norm $\|f\| = \sqrt{\langle f, f \rangle}$.

A *functional time series* is a sequence of $\mathcal{H}$-valued random variables $\{X_t\}_{t \in \mathbb{Z}}$, where each X_t is a random element of $\mathcal{H}$, i.e.,

$$X_t : \Omega \to \mathcal{H}, \quad t \in \mathbb{Z}.$$

A functional time series $\{X_t\}_{t\in\mathbb{Z}}$ is said to be **Mean-square continuous** if $\mathbb{E}\|X_t\|^2 < \infty$ for all t; **Second-order stationary** if the mean function $\mu(t) := \mathbb{E}[X_t]$ is constant over time and the autocovariance operator

$$\Gamma_h = \mathrm{Cov}(X_{t+h}, X_t) = \mathbb{E}[(X_{t+h} - \mu) \otimes (X_t - \mu)]$$

depends only on the lag h, where $\otimes$ denotes the tensor product.

3 Revisão de literatura

The work of Aue, Norinho e Hörmann (2015) proposes a simplification of the functional time series prediction problem by reducing it to a multivariate forecasting problem, thereby allowing the use of well-established tools, as opposed to the methodology initially developed in Bosq (2000). The proposed algorithm consists of three steps: first, a number d of principal components is selected to retain $(\alpha \cdot 100)\%$ of the variance of the original data; then, given a forecast horizon h, a VAR(p) model is fitted to the principal components, and an h-step-ahead forecast is computed; finally, the multivariate forecasts are transformed back to the original functional space via a truncated Karhunen–Loève representation. It is also shown that the one-step-ahead forecast from a VAR(1) model in the second step is asymptotically equivalent to that of a FAR(1) model, which simplifies the forecasting task. Another important contribution of the paper is the proposal of a fully automatic and joint procedure for selecting the model order p and the number of components d through the minimization of a functional final prediction error (fFPE) criterion given by

$$\mathit{fFPE}(p, d) = \frac{n + pd}{n - pd}\mathrm{tr}(\hat{\Sigma}_Z) + \sum_{l>d} \hat{\lambda}_l, \tag{1}$$

which makes the proposed methodology entirely data-driven. The possibility of including exogenous variables in the model is also supported without major theoretical complications. Finally, simulation studies and applications to real data compare the performance of the new methodology with that of Hyndman e Ullah (2007), which carries out forecasting by treating the principal component scores as univariate time series, and Bosq (2000), using the autoregressive order selection criterion proposed by Kokoszka e Reimherr (2013). In both settings, the new method outperformed the alternatives.

Trabalhos relacionados à estimação de séries temporais funcionais.

Previsão.

Tentativas de prever funções de densidade de probabilidade.

Artigo Horta.

Introdução a FDA: Dabo-Niang e Frévent (2024).

Aue, Norinho e Hörmann (2015),Bathia, Yao e Ziegelmann (2010), Benko, Härdle e Kneip (2009), Besse e Ramsay (1986), Bosq (2000), Dabo-Niang et al. (2008), Dauxois, Pousse e Romain

(1982), Ferraty e Vieu (2003), Hall e Vial (2006), Ferraty (2006), Horta e Ziegelmann (2018), Hron et al. (2016), Müller, Dacorogna e Pictet (1998), Petersen e Müller (2016), Ramsay e Dalzell (1991), Ramsay e Silverman (2002), Ramsay e Silverman (2005), Ramsay, Hooker e Graves (2009)

4 Metodologia

Theorem 1. *This is the first theorem.*

Lemma 1. *This is a lemma that follows the theorem.*

Definition 1. *This is a definition related to the previous results.*

Definition 2. *In the theory of random processes, a sequence* $\{X_n\}_{n=1}^{\infty}$ *is said to be* ψ*-mixing if the dependence between past and future events decreases as they become further apart in time, according to a specific mixing coefficient.*

Let $\{X_n\}_{n=1}^{\infty}$ *be a sequence of random variables defined on a probability space* $(\Omega, \mathcal{F}, P)$*. The sequence is called* ψ-mixing *if there exists a function* $\psi(n)$ *such that for any two* σ*-algebras* $\mathcal{F}_a^b = \sigma(X_a, X_{a+1}, \ldots, X_b)$ *and* $\mathcal{F}_c^d = \sigma(X_c, X_{c+1}, \ldots, X_d)$ *with* $a \leq b < c \leq d$*, the following holds:*

$$\psi(n) = \sup_{A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+n}^{\infty}} |P(A \cap B) - P(A)P(B)|,$$

where $\psi(n) \to 0$ *as* $n \to \infty$*.*

The sequence is said to be ψ*-mixing if* $\psi(n) \to 0$ *as* $n \to \infty$*. This condition implies that the events in the distant past and the far future become asymptotically independent.*

Definition 3. *Let* $\mathcal{X}$ *be a domain and* $h_0(x)$ *a reference probability density function on* $\mathcal{X}$*. The* Bayes space $B^2(\mathcal{X}, h_0)$ *is defined as the space of all functions* $h(x) > 0$ *such that:*

$$\log \frac{h(x)}{h_0(x)} \in L^2(\mathcal{X}),$$

where $L^2(\mathcal{X})$ *denotes the space of square-integrable functions on* $\mathcal{X}$*. The inner product between two elements* $h_1(x), h_2(x) \in B^2(\mathcal{X}, h_0)$ *is given by:*

$$\langle h_1, h_2 \rangle_{B^2} = \int_{\mathcal{X}} \log \frac{h_1(x)}{h_0(x)} \log \frac{h_2(x)}{h_0(x)} h_0(x) dx.$$

The associated norm is:

$$\|h\|_{B^2} = \left(\int_{\mathcal{X}} \left(\log \frac{h(x)}{h_0(x)} \right)^2 h_0(x) dx \right)^{\frac{1}{2}}.$$

4.1 Estimação de densidade

4.2 Dados

4.2.1 Análise exploratória

5 Resultados

6 Considerações finais

Referências

AUE, A.; NORINHO, D. D.; HÖRMANN, S. On the prediction of stationary functional time series. *Journal of the American Statistical Association*, Taylor & Francis, v. 110, n. 509, p. 378–392, 2015.

BATHIA, N.; YAO, Q.; ZIEGELMANN, F. Identifying the finite dimensionality of curve time series. 2010.

BENKO, M.; HÄRDLE, W.; KNEIP, A. Common functional principal components. 2009.

BESSE, P.; RAMSAY, J. O. Principal components analysis of sampled functions. *Psychometrika*, Springer-Verlag, v. 51, n. 2, p. 285–311, 1986.

BOSQ, D. *Linear processes in function spaces: theory and applications*. [S.l.]: Springer Science & Business Media, 2000. v. 149.

DABO-NIANG, S. et al. Functional linear regression with functional response: application to prediction of electricity consumption. In: SPRINGER. *Functional and Operatorial Statistics*. [S.l.], 2008. p. 23–29.

DABO-NIANG, S.; FRÉVENT, C. Uncovering data across continua: An introduction to functional data analysis. *BRIN MATHEMATICS RESEARCH CENTER*, v. 71, n. 7, p. 869, 2024.

DAUXOIS, J.; POUSSE, A.; ROMAIN, Y. Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of multivariate analysis*, Elsevier, v. 12, n. 1, p. 136–154, 1982.

FERRATY, F. *Nonparametric functional data analysis*. [S.l.]: Springer, 2006.

FERRATY, F.; VIEU, P. Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis*, Elsevier, v. 44, n. 1-2, p. 161–173, 2003.

GERTHEISS, J. et al. Functional data analysis: An introduction and recent developments. *Biometrical Journal*, v. 66, n. 7, p. e202300363, 2024. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.202300363>.

HALL, P.; VIAL, C. Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Oxford University Press, v. 68, n. 4, p. 689–705, 2006.

HORTA, E.; ZIEGELMANN, F. Dynamics of financial returns densities: A functional approach applied to the bovespa intraday index. *International Journal of Forecasting*, Elsevier, v. 34, n. 1, p. 75–88, 2018.

HRON, K. et al. Simplicial principal component analysis for density functions in bayes spaces. *Computational Statistics & Data Analysis*, Elsevier, v. 94, p. 330–350, 2016.

HYNDMAN, R. J.; ULLAH, M. S. Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, Elsevier, v. 51, n. 10, p. 4942–4956, 2007.

KOKOSZKA, P.; REIMHERR, M. Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, Wiley Online Library, v. 34, n. 1, p. 116–129, 2013.

MÜLLER, U. A.; DACOROGNA, M. M.; PICTET, O. V. Heavy tails in high-frequency financial data. *A practical guide to heavy tails: Statistical techniques and applications*, Boston, MA: Birkhäuser Boston, Inc, p. 55–78, 1998.

PETERSEN, A.; MÜLLER, H.-G. Functional data analysis for density functions by transformation to a hilbert space. 2016.

RAMSAY, J.; SILVERMAN, B. *Functional Data Analysis*. Springer, 2005. (Springer Series in Statistics). ISBN 9780387400808. Disponível em: <https://books.google.com.br/books?id=mU3dop5wY_4C>.

RAMSAY, J. O.; DALZELL, C. Some tools for functional data analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Oxford University Press, v. 53, n. 3, p. 539–561, 1991.

RAMSAY, J. O.; HOOKER, G.; GRAVES, S. *Functional Data Analysis with R and MATLAB*. 1st. ed. [S.l.]: Springer Publishing Company, Incorporated, 2009. ISBN 0387981845.

RAMSAY, J. O.; SILVERMAN, B. W. *Applied functional data analysis: methods and case studies*. [S.l.]: Springer, 2002.

7 Apêndice

7.1 Figuras

7.2 Tabelas