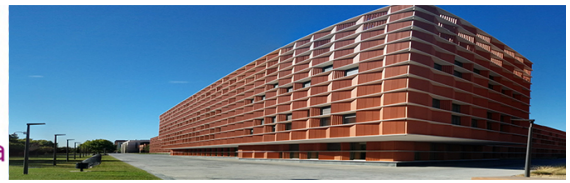
Jornada de presentación de los miembros del Instituto UC3M-BS on Financial Big Data

Pedro Galeano
pedro.galeano@uc3m.es

Departament of Statistics
Institute UC3M-BS on Financial Big Data

Universidad Carlos III de Madrid
www.uc3m.es

UC3M-BS Institute of Financial Big Data



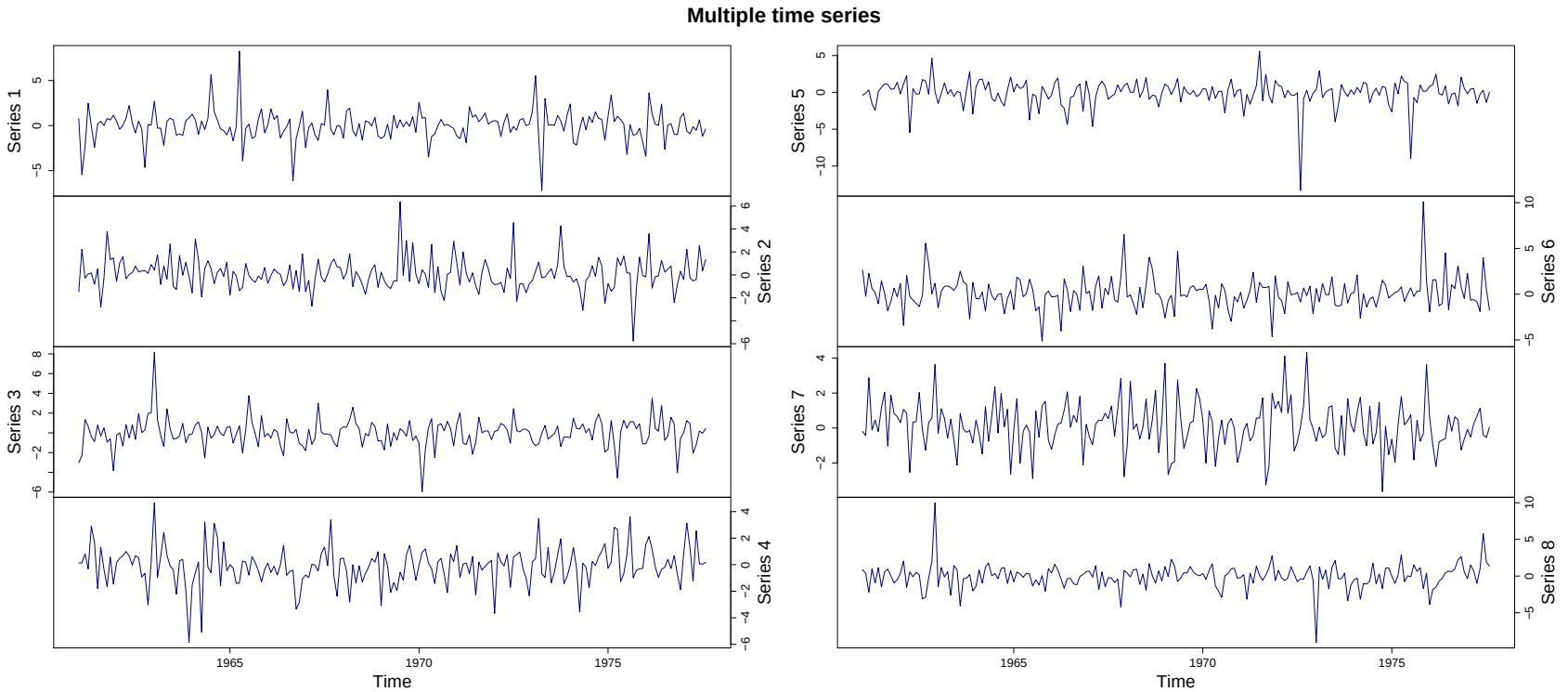
1. Current and past research interests
   - Time series analysis
   - Functional data analysis

2. Research interests for the future
   - Sparse modeling for high-dimensional temporal data
   - Statistical analysis of data from large networks

1. Current and past research interests
   - Time series analysis
   - Functional data analysis
2. Research interests for the future
   - Sparse modeling for high-dimensional temporal data
   - Statistical analysis of data from large networks

Time series analysis

- **Time series** are sequences of data points consisting of consecutive measures taken over a finite time interval.
- **Two main goals**: to understand the time and cross dependency of the time series and to obtain reliable forecasts.



Time series analysis: Outlier detection

- The presence of **outliers** in time series can lead to model misspecification, biased parameter estimation, and poor forecasts.
- This kind of data is usually corrected by means of methods that are based on identifying these anomalous observations.
- **Contributions** (with D. Peña and R. S. Tsay):
  - Procedures for detecting outliers in univariate time series by means of model selection criteria; and
  - Procedures for detecting outliers in multiple time series by means of projection pursuit methods.

Time series analysis: Changepoint analysis

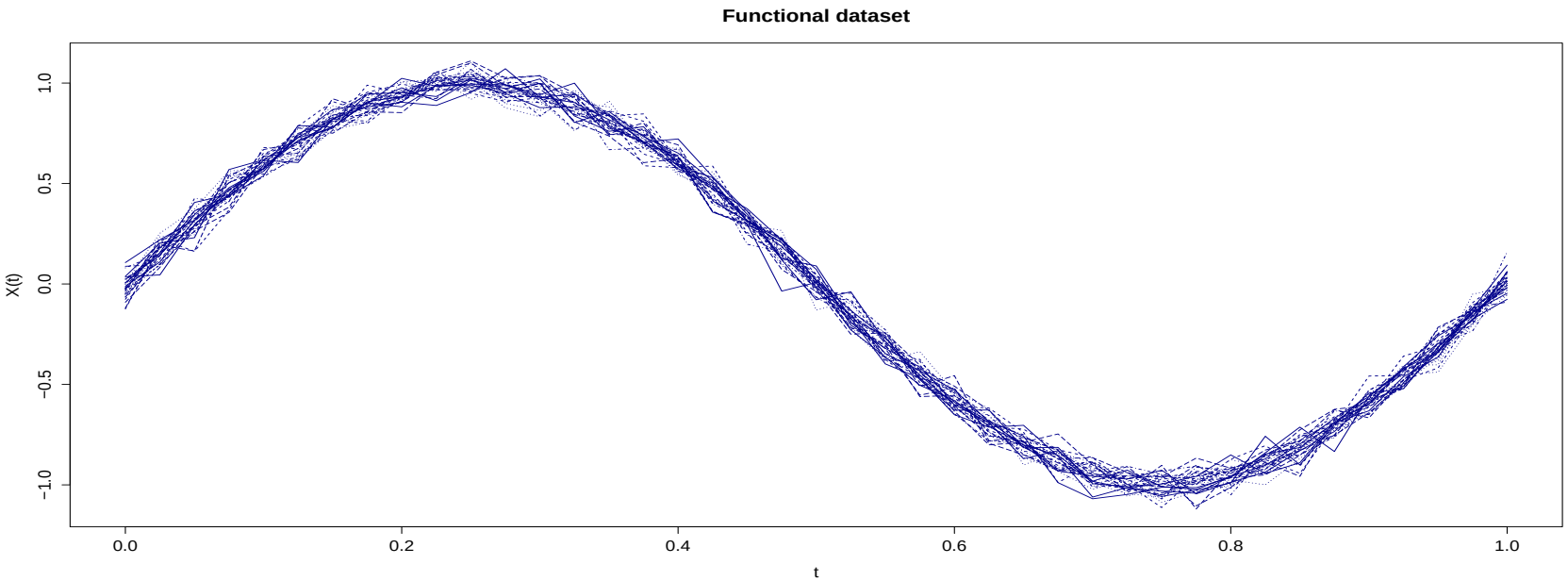
- **Changepoint analysis** or **time series segmentation** deals with dividing a sequence of observations into discrete segments that contains observations with common characteristics.
- The goal is to detect the number of segments and to estimate the piecewise common characteristics.
- **Contributions** (with D. Peña, D. Wied and K. Pape):
  - Procedures for detecting change points in the covariances as well the correlations of time series;
  - On-line monitoring procedures for similar issues; and
  - Methods for detecting change points in the rate parameter of Poisson processes.

Time series analysis: Financial time series

- **Financial time series analysis** is concerned with the theory and practice of asset valuation over time.
- Particularly, financial returns have certain characteristics that are not shared by time series in other fields.
- One of the major challenges is to provide reliable estimates of pasts volatilities and to give good forecasts of future volatilities.
- **Contributions** (with R. S. Tsay, M. C. Ausin, A. Virbickaite, H. Lopes and P. Ghosh):
  - Procedures for detecting change points in volatilities;
  - Bayesian estimation methods for model fitting; and
  - Bayesian methods for selecting optimal portfolios and for estimating the Value at Risk (VaR) of these portfolios.

Functional data analysis

- **Functional data analysis** provides information about curves, surfaces or anything else varying over a continuum (time, spatial location, wavelength, probability, etc...).
- **Many different goals**: Dimension reduction, supervised and unsupervised classification, outlier detection, forecasting...



Functional data analysis: Distance based methods

- **Distances between observations** are key tools for developing methods for data analysis.
- For instance, distance methods exist for supervised and unsupervised classification, outlier detection and model selection, among many others.
- **Contributions** (with R. E. Lillo and E. Joseph):
  - The Mahalanobis distance for functional observations;
  - Supervised classification methods for functional observations based on the functional Mahalanobis distance; and
  - Two-sample Hotelling's T^2 statistics based on the functional Mahalanobis distance.

Functional data analysis: Depth based methods

- **Depth functions** provides an order of functional observations based on their departure from the centre.
- Therefore, depth functions measure the degree of outlyingness of a given point with respect to a functional data cloud or its underlying distribution.
- **Contributions** (with M. Febrero-Bande, W. González-Manteiga, R. E. Lillo and C. Sguera):
  - Depth based methods for functional outlier detection; and
  - Depth based methods for functional supervised classification.

Functional data analysis: Functional regression models

- **Regression methods** are frequently used in data analysis for predicting one or several variables of interest in terms of others.
- In the functional framework there are models and inferential methods in the situation in which the predictors and/or the responses have a functional nature.
- **Contributions** (with M. Febrero-Bande and W. González-Manteiga):
  - Influence measures for functional linear regression models; and
  - Analysis of PCA and PLS estimators of functional regression models.

1. Current and past research interests
   - Time series analysis
   - Functional data analysis

2. Research interests for the future
   - Sparse modeling for high-dimensional temporal data
   - Statistical analysis of data from large networks

Sparse modeling for high-dimensional temporal data

- A **sparse statistical model** is one having only a small number of parameters (or predictors).
- Sparse methods have been recently developed with two main goals:
  - to accurately predict future high-dimensional observations; and
  - to gain insight into the relationship between many different features.
- Most of the work has been mainly placed in regression models but also on principal component analysis, supervised and unsupervised classification and canonical correlation analysis, among others.
- **Goal**: Sparse modeling for high-dimensional temporal data (time series and functional data).

- **Statistical analysis of network data** consists on the analysis of measurements that are either of or from a system conceptualized as a network.
- **Current project** with Bank of Santander regarding several aspects of the analysis of their network of clients.
- **Characteristics**: large, dynamic, heterogeneous, noisy,...
- **Goals**: develop new methods for:
  - describing network characteristics (dynamic or not);
  - sampling and estimation in networks (dynamic or not);
  - modeling and prediction for processes on networks (dynamic or not);
  - ...

1. Current and past research interests
   - Time series analysis
   - Functional data analysis

2. Research interests for the future
   - Sparse modeling for high-dimensional temporal data
   - Statistical analysis of data from large networks