Functional Data Analysis of High-Frequency Household Energy Consumption Curves for Policy Evaluation

Simone Vantini

MOX - Department of Mathematics, Politecnico di Milano, Milano, Italy.
From 1D-data to Functional data

- Functional Data
  - Functions (i.e., curves, surfaces, trajectories, BW and RGB images, ...)

- High-dimensional Data
  - "Very long" vectors

- Multivariate Data
  - Vectors

- Univariate Data
  - Numbers
Laser Welding of Metal Sheets

Functional Two-population Test: a Simple Problem with Complex Data
Functional Two-population Test: 
a Simple Problem with Complex Data

\[ y_{1i} \sim i.i.d. \ Y_1 \quad i = 1, \ldots, n_1 \]
\[ y_{2i} \sim i.i.d. \ Y_2 \quad i = 1, \ldots, n_2 \]

where \( Y_1 \) and \( Y_2 \) are \( L^2 \)-valued random functions

Functional \( t \)-test:

\[ H_0 : \ Y_1 = Y_2 \]
\[ H_1 : \ Y_1 \neq Y_2 \]
Approaches to Null Hypothesis Significance Testing in FDA

Functional data are not Gaussian
In FDA sample sizes are intrinsically small

Rejections are required to be located along the domain

Null Hypothesis Testing for FD
- Parametric
- Asymptotic
- Nonparametric

Global
Local
Research question:

Can we monitor the gap between the two sheets despite of the nuisance effect due to location by looking at the backward spectrum?
We want to test:

- **If and in which wavelength intervals** the **gap** effect is significant
- **If and in which wavelength intervals** the **location** effect is significant

\[ y_{ijl}(\lambda) = \mu(\lambda) + \alpha_i(\lambda) + \beta_j(\lambda) + \varepsilon_{ijl}(\lambda) \]

- \( \mu(\lambda) \): common mean
- \( \alpha_i(\lambda) \): gap effect \((i=1,2,3)\)
- \( \beta_j(\lambda) \): location effect \((j=1,\ldots,5)\)
- \( \varepsilon_{ijl}(\lambda) \): independent errors \((l=1,2,3)\)
Functional F-tests

**P-value functions**

**Significant regions**

**Gap**

**Location**
Research question:
Do household electric appliances have a specific load signature that can be identified over time in the various consumption curves?

<table>
<thead>
<tr>
<th>Data Family</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative Data</td>
<td>Client ID</td>
</tr>
<tr>
<td></td>
<td>Display delivery date</td>
</tr>
<tr>
<td></td>
<td>Municipality</td>
</tr>
<tr>
<td></td>
<td>Display Version</td>
</tr>
<tr>
<td></td>
<td>Presence of microgeneration devices</td>
</tr>
<tr>
<td>Profilation Data (425 Households)</td>
<td>Contractual power</td>
</tr>
<tr>
<td></td>
<td>Number of People in the household</td>
</tr>
<tr>
<td></td>
<td>Age and Sex</td>
</tr>
<tr>
<td></td>
<td>Electric Appliances owned (34 items)</td>
</tr>
<tr>
<td>Load Curves (1064 Households, ~600 days)</td>
<td>Size of the household</td>
</tr>
<tr>
<td></td>
<td>Sampled every 15 minutes</td>
</tr>
</tbody>
</table>
Electricity Consumption $= f(\text{Appliances}) + \epsilon$
Functional-on-scalar Regression Model

\[ f = \epsilon(t) \]
Functional-on-scalar Regression Model

\[ y(t)_f = \beta_{Baseline}(t) + \beta_{Hi.tech}(t)I_{fH} + \beta_{Lo.tech}(t)I_{fL} + \epsilon(t)_f \]
Building the Model Response:
Electricity Consumption Curve Smoothing

Fourier smoothing
of the average consumption curve for each household
Building the Model Regressors: Appliance Clustering

- **Basic Appliances** – **Blue**
- **Low Tech/High Consumption Appliances** – **Yellow**
- **High Tech Appliances** – **Red**
Define a ownership index for each cluster $k$ and each family $f$ as

$$I_{fk} = \frac{1}{N_k} \sum_{i=1}^{N_K} a_{fik}, \quad \forall k \in \{B, H, L\}$$
Functional t-tests

Baseline

High Tech

Low Tech

$\beta(t)$

$P(t)$
Energy Disaggregation
From Consumptions to Consumption Variations

Functional-on-scalar Linear Model (Time Derivatives)

\[ D_y(t)_f = \beta_{Baseline}(t) + \beta_{Hi.tech}(t)I_{fH} + \beta_{Lo.tech}(t)I_{fL} + \epsilon(t)_f \]

Consumption curves

Consumption variation curves
Functional t-tests

Baseline

High Tech

Low Tech

$\beta(t)$

$P(t)$
Many Other Applications of FDA
Some References

Background:

General Methodology:

Specific Applications:

Software: