Classification of Nonparametric Time Trends

Michael Vogt
University of Bonn

Joint work with Oliver Linton

CIRM
February 18, 2016

Introduction

- In many applications, we observe a multitude of time series $\mathcal{Y}_i = \{Y_{it} : 1 \le t \le T\}$ with $1 \le i \le n$.
- The observed time series often exhibit a nonstationary behaviour. In particular, their stochastic behaviour often appears to gradually change over time.
- Processes with time-varying parameters, or more generally, locally stationary processes provide a neat way to model such a behaviour. Simple examples are

$$\begin{aligned} \text{Trend model:} \quad & Y_{it} = m_i\big(\tfrac{t}{T}\big) + \varepsilon_{it}. \\ \text{Volatility model:} \quad & Y_{it} = \sigma_i\big(\tfrac{t}{T}\big)\,\varepsilon_{it}. \\ \text{AR model:} \quad & Y_{it} = a_i\big(\tfrac{t}{T}\big) Y_{it-1} + \varepsilon_{it}. \end{aligned}$$

Introduction

- In most applications, it is very restrictive to assume that the parameter functions are the same for all time series.
- However, it is often natural to impose a group structure on the time series: we may suppose that the time series can be grouped into a number of classes whose members share the same parameter functions.
- In the talk, we are interested in the statistical question how to estimate the unknown group structure from the data.

Model setting

Data: We observe n different time series

$$\mathcal{Y}_i = \{Y_{it} : 1 \leq t \leq T\}$$

with $1 \leq i \leq n$. Here, $T \to \infty$, whereas n may either be bounded or $n \to \infty$.

Time trend model: Each time series $\mathcal{Y}_i$ follows the model

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} \quad \text{for } 1 \leq t \leq T,$$

where m_i are unknown nonparametric trend functions.

Error structure: We restrict attention to the simple case that ε_{it} is i.i.d. both across i and t with $\mathbb{E}[\varepsilon_{it}] = 0$.

Model setting

Group structure: There are K groups of time series $G_1, \ldots, G_K$ with $G_1 \dot{\cup} \ldots \dot{\cup} G_K = \{1, \ldots, n\}$ s.t. for each $k \in \{1, \ldots, K\}$,

$$m_i = m_j \quad \text{for all } i, j \in G_k.$$

Hence, the members of the class G_k all have the same time trend function.

Aim: We want to estimate the unknown groups $G_1, \ldots, G_K$ along with their unknown number K.

Estimation of the classes $G_1, \ldots, G_K$

Define the squared L_2-distance between m_i and m_j by

$$\Delta_{ij} = \int (m_i(w) - m_j(w))^2 \pi(w) dw,$$

where π is some weight function. Estimate this by

$$\widehat{\Delta}_{ij} = \int \left(\widehat{m}_i(w) - \widehat{m}_j(w)\right)^2 \pi(w) dw,$$

where $\widehat{m}_i$ is a standard NW estimator of the form

$$\widehat{m}_i(w) = \frac{\sum_{t=1}^T W_h(\frac{t}{T} - w) Y_{it}}{\sum_{t=1}^T W_h(\frac{t}{T} - w)}.$$

Here, h is the bandwidth and W is a kernel with $W_h(x) = h^{-1} W(x/h)$.

Estimation of the classes $G_1, \dots, G_K$

Preliminary estimation problem:

- Pick some time series i and let $G \in \{G_1, \dots, G_K\}$ be the unknown class to which i belongs.
- Let $S \subseteq \{1, \dots, n\}$ be some index set with $G \subseteq S$.
- We want to estimate the group G from the set S.

Notation: Denote the ordered distances by

$$\Delta_{i(1)} \le \Delta_{i(2)} \le \dots \le \Delta_{i(n_S)}$$
$$\widehat{\Delta}_{i[1]} \le \widehat{\Delta}_{i[2]} \le \dots \le \widehat{\Delta}_{i[n_S]}$$

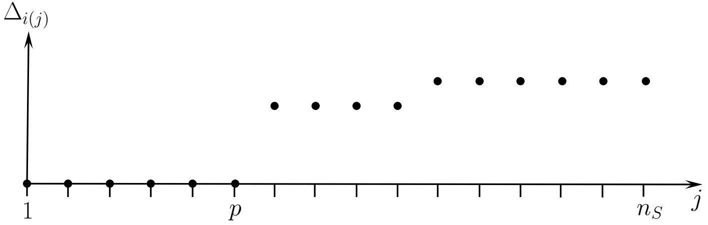
with $n_S = |S|$.

Estimation of the classes $G_1, \ldots, G_K$

The ordered distances $\Delta_{i(j)}$ have the following property: There exists a point $p = p_{i,S}$ such that

$$\Delta_{i(j)} \begin{cases} = 0 & \text{for } j \leq p \\ \geq c & \text{for } j > p \end{cases}$$

with $c = \Delta_{i(p+1)} > 0$. As a consequence, $G = \{(1), \ldots, (p)\}$.

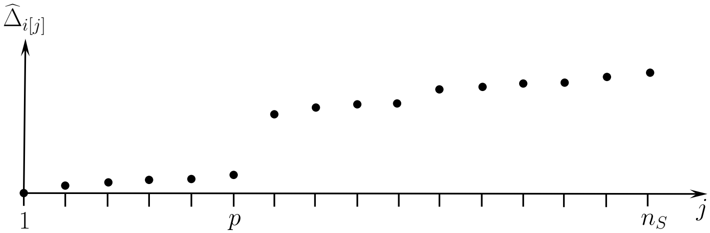


Estimation of the classes $G_1, \dots, G_K$

Under appropriate regularity conditions, it holds that

$$\widehat{\Delta}_{i[j]} \begin{cases} = o_p(1) & \text{for } j \leq p \\ \geq c + o_p(1) & \text{for } j > p \end{cases}$$

with some $c > 0$. If $p = |G|$ were known, we could thus simply estimate $G = \{(1), \dots, (p)\}$ by $\widetilde{G} = \{[1], \dots, [p]\}$.

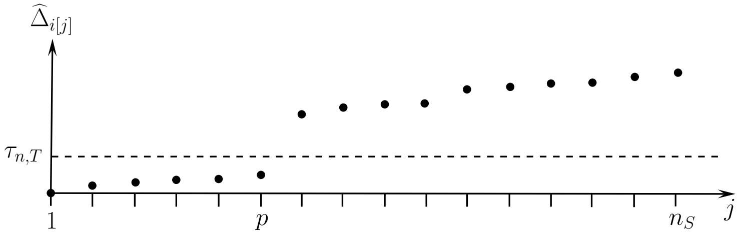


Estimation of the classes $G_1, \ldots, G_K$

Under appropriate regularity conditions, it holds that

$$\widehat{\Delta}_{i[j]} \begin{cases} = o_p(1) & \text{for } j \leq p \\ \geq c + o_p(1) & \text{for } j > p \end{cases}$$

with some $c > 0$. If $p = |G|$ were known, we could thus simply estimate $G = \{(1), \ldots, (p)\}$ by $\widetilde{G} = \{[1], \ldots, [p]\}$.

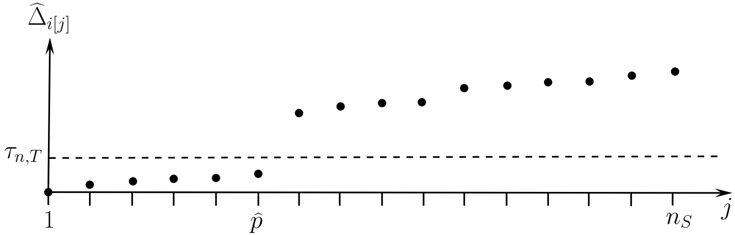


Estimation of the classes $G_1, \dots, G_K$

As p is not known, we estimate it by a thresholding procedure: Let $\tau_{n,T} \searrow 0$ such that $\max_{1 \le j \le p} \widehat{\Delta}_{i[j]} \le \tau_{n,T}$ with prob. tending to 1 and estimate $p = p_{i,S}$ by

$$\widehat{p} = \widehat{p}_{i,S} = \max\left\{ j \in \{1, \dots, n_S\} : \widehat{\Delta}_{i[j]} \le \tau_{n,T} \right\}.$$

Our estimator of G is now defined as $\widehat{G} = \{[1], \dots, [\widehat{p}]\}$.



Estimation of the classes $G_1, \ldots, G_K$

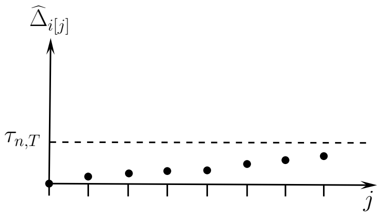
Iterative algorithm:

1[st] Step:
- Set $S_1 = \{1, \ldots, n\}$, pick some index $i_1 \in S_1$, and write $\widehat{\Delta}_{i_1[1]} \leq \ldots \leq \widehat{\Delta}_{i_1[n_{S_1}]}$.
- Compute $\widehat{p} = \widehat{p}_{i_1, S_1}$ and estimate the class to which i_1 belongs by $\widehat{G}_1 = \{[1], \ldots, [\widehat{p}]\}$.

k^{th} Step:
- Let $\widehat{G}_1, \ldots, \widehat{G}_{k-1}$ be the class estimates from the previous iteration steps.
- Set $S_k = \{1, \ldots, n\} \setminus \bigcup_{\ell=1}^{k-1} \widehat{G}_\ell$, pick some index $i_k \in S_k$, and write $\widehat{\Delta}_{i_k[1]} \leq \ldots \leq \widehat{\Delta}_{i_k[n_{S_k}]}$.
- Compute $\widehat{p} = \widehat{p}_{i_k, S_k}$ and estimate the class to which i_k belongs by $\widehat{G}_k = \{[1], \ldots, [\widehat{p}]\}$.

Estimation of the classes $G_1, \dots, G_K$

We iterate the algorithm $\widehat{K}$ times until $\widehat{\Delta}_{i_{\widehat{K}}[j]} \le \tau_{n,T}$ for all $1 \le j \le n_{S_{\widehat{K}}}$. In this case, $S_{\widehat{K}}$ is not split into two parts any more and $\widehat{G}_{\widehat{K}} = S_{\widehat{K}}$.

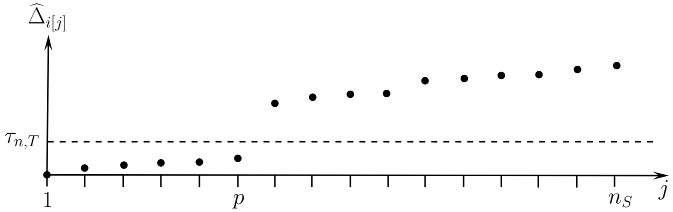


Estimators:

- The algorithm produces the partition $\{\widehat{G}_k : 1 \le k \le \widehat{K}\}$, which serves as our estimator of the class structure $\{G_k : 1 \le k \le K\}$.
- The number of classes K is implicitly estimated by the number of iterations $\widehat{K}$.

Choice of the threshold parameter $\tau_{n,T}$

- Let $i \in G$ and suppose we want to estimate the unknown class G.
- As discussed above, we would ideally like to choose $\tau_{n,T}$ s.t. $\max_{1 \le j \le p} \widehat{\Delta}_{i[j]} \le \tau_{n,T}$.



- As $\max_{1 \le j \le p} \widehat{\Delta}_{i[j]} = \max_{j \in G} \widehat{\Delta}_{ij}$ with prob. tending to one, this means that we would like to choose $\tau_{n,T}$ s.t.

$$\max_{j \in G} \widehat{\Delta}_{ij} \le \tau_{n,T}.$$

Choice of the threshold parameter $\tau_{n,T}$

- One can show that for any $j \in G$ with $j \neq i$,

$$Th^{1/2}\widehat{\Delta}_{ij} - h^{-1/2}\mathcal{B} \xrightarrow{d} N(0, \mathcal{V}), \qquad (*)$$

where

$$\mathcal{B} = 2\,\sigma^2 \|W\|^2 \int \pi(x)dx$$
$$\mathcal{V} = 8\,\sigma^4 \|W * W\|^2 \int \pi^2(x)dx$$

and $\sigma^2 = \mathbb{E}[\varepsilon_{it}^2]$. Moreover,

$$\|W\|^2 = \int W^2(x)dx$$
$$\|W * W\|^2 = \int \Big(\int W(x)W(x+y)dx \Big)^2 dy.$$

Choice of the threshold parameter $\tau_{n,T}$

- Roughly speaking, $(*)$ says that

$$\widehat{\Delta}_{ij} \approx \Delta^*_{ij} := \frac{\mathcal{B}}{Th} + \frac{\sqrt{\mathcal{V}}}{Th^{1/2}} Z_{ij} \quad \text{with} \quad Z_{ij} \sim N(0,1). \qquad (**)$$

- Neglecting the approximation error in $(**)$, we want to choose $\tau_{n,T}$ s.t. $\max_{j \in G} \Delta^*_{ij} \le \tau_{n,T}$. (Here, we set $\Delta^*_{ii} = 0$ since $\widehat{\Delta}_{ii} = 0$ by construction.) We have

$$\max_{j \in G} \Delta^*_{ij} = \max_{j \in G_{-i}} \Delta^*_{ij} = \frac{\mathcal{B}}{Th} + \frac{\sqrt{\mathcal{V}}}{Th^{1/2}} \max_{j \in G_{-i}} Z_{ij}$$

with $G_{-i} = G \setminus \{i\}$.

Choice of the threshold parameter $\tau_{n,T}$

- Since the variables Z_{ij} are standard normal,

$$\mathbb{P}\Big(\max_{j \in G_{-i}} Z_{ij} \geq (2 \log |G|)^{1/2}\Big) \leq \frac{1}{\sqrt{4\pi \log |G|}}.$$

- Hence,

$$\max_{j \in G} \Delta^*_{ij} \leq \frac{\mathcal{B}}{Th} + \frac{\sqrt{\mathcal{V}}}{Th^{1/2}} (2 \log |G|)^{1/2}$$

  with prob. approaching 1 as $|G| \to \infty$.

- This suggests that an appropriate threshold level is given by

$$\tau_{n,T} = \frac{\mathcal{B}}{Th} + \frac{\sqrt{\mathcal{V}}}{Th^{1/2}} (2 \log |G|)^{1/2}.$$

Choice of the threshold parameter $\tau_{n,T}$

- Since the variables Z_{ij} are standard normal,

$$\mathbb{P}\Big(\max_{j\in G_{-i}} Z_{ij} \geq (2\log|G|)^{1/2}\Big) \leq \frac{1}{\sqrt{4\pi\log|G|}}.$$

- Hence,

$$\max_{j\in G} \Delta^*_{ij} \leq \frac{\mathcal{B}}{Th} + \frac{\sqrt{\mathcal{V}}}{Th^{1/2}}(2\log|G|)^{1/2}$$

with prob. approaching 1 as $|G| \to \infty$.

- This suggests that an appropriate threshold level is given by

$$\tau_{n,T} = \frac{\widehat{\mathcal{B}}}{Th} + \frac{\sqrt{\widehat{\mathcal{V}}}}{Th^{1/2}}(2\log n)^{1/2}.$$

Theoretical results

Consistency of the class estimates $\{\widehat{G}_k : 1 \leq k \leq \widehat{K}\}$:

Let the threshold parameter $\tau_{n,T}$ converge to zero such that for $1 \leq k \leq K$,

$$\mathbb{P}\Big(\max_{i,j \in G_k} \widehat{\Delta}_{ij} \leq \tau_{n,T} \Big) \to 1.$$

Then under appropriate regularity conditions,

$$\mathbb{P}(\widehat{K} \neq K) = o(1)$$

and

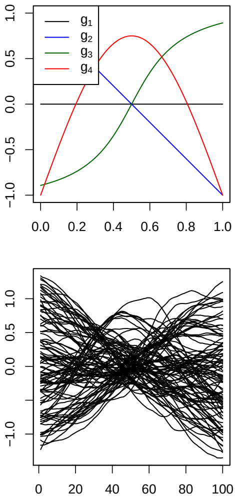
$$\mathbb{P}\Big(\big\{ \widehat{G}_k : 1 \leq k \leq \widehat{K} \big\} \neq \big\{ G_k : 1 \leq k \leq K \big\} \Big) = o(1).$$

Illustrative example

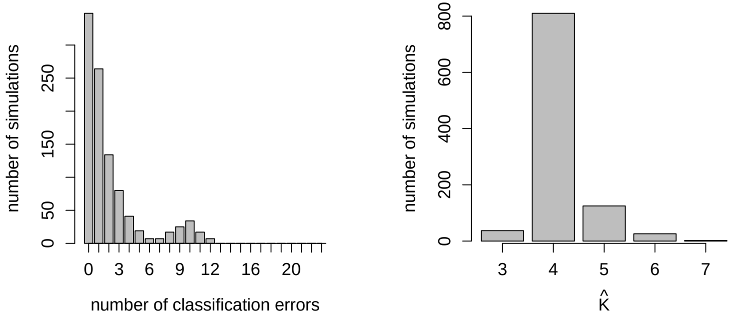
We consider a simulation design with $T = 100$, $n = 100$, and the four groups

$$G_1 = \{1, \ldots, 40\}$$
$$G_2 = \{41, \ldots, 70\}$$
$$G_3 = \{71, \ldots, 90\}$$
$$G_4 = \{91, \ldots, 100\}.$$

The error variance $\mathbb{E}[\varepsilon_{it}^2]$ is equal to 1.



Illustrative example



Relationship to functional data clustering

A functional data model:

- $Y_{it} = m_i(\frac{t}{T}) + \varepsilon_{it}$ for $1 \leq t \leq T$ and $1 \leq i \leq n$ with i.i.d. noise ε_{it}.
- The curves $m_i = (m_i(w))_{w \in [0,1]}$ are Gaussian processes.
- There are clusters of indices $G_1, \ldots, G_K$ s.t. the Gaussian processes m_i have the same mean and covariance structure within each cluster.

Relationship to our model:

- The curves m_i in the above functional data model are random. Within each group, the observed sample paths m_i are realizations from the same Gaussian process.
- In our setting, the curves m_i are deterministic. Within each group, the curves m_i are exactly the same.

Literature

Abraham, C., Cornillon, P. A., Matzner-Løber, E. & Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics.*

James, M. & Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association.*

Vogt, M. & Linton, O. (2015). Classification of nonparametric regression functions in longitudinal data models. Forthcoming in *Journal of the Royal Statistical Society: Series B*.

Degras, D., Xu, Z., Zhang, T. & Wu, W. B. (2012). Testing for Parallelism Among Trends in Multiple Time Series. *IEEE Transactions on Signal Processing.*