Multiscale inference for nonparametric time trends

Marina Khismatullina Michael Vogt

August 24, 2021

HCM Symposium

Table of contents

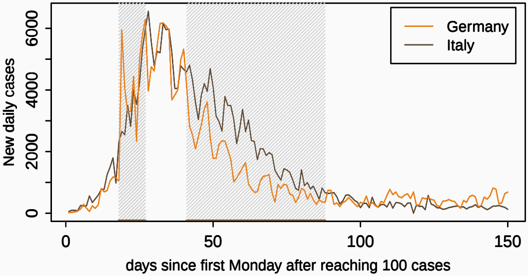
1. Introduction
2. Model
3. Testing procedure
4. Theoretical properties
5. Application

Introduction

Motivation

Aim of the paper

To develop new inference methods that allow to *identify* and *locate* differences between nonparametric trend curves with dependent errors.



Research question: Out of many given intervals, how to find those where the trends are significantly different?

Motivation

Why is it relevant?

Finding systematic differences between trends = basis for further research

$\Rightarrow$ understanding which government policies are more effective.

Why is it difficult?

Testing many hypotheses at the same time = multiple testing problem

$\Rightarrow$ large probability of one true null hypothesis being rejected.

Is it limited to COVID-19?

No! Our method = general method for comparing nonparametric trends

$\Rightarrow$ new statistical test for equality of nonparametric trend curves.

Literature

Comparison of deterministic trends:

- Park et al. (2009), Degras et al. (2012), Zhang et al. (2012), Hidalgo and Lee (2014), Chen and Wu (2019).

Multiscale tests:

- Chaudhuri and Marron (1999, 2000), Hall and Heckman (2000), Dümbgen and Spokoiny (2001), Park et al. (2009).

Studies of COVID-19:

- Dong et al. (2020), Gu et al. (2020), Li and Linton (2020), Jiang et al. (2020).

Model

Motivation for the model

We observe n time series $\mathcal{X}_i = \{X_{it} : 1 \leq t \leq T\}$ of length T.

X_{it} are non-negative integers $\Rightarrow$ can be modelled by a Poisson distribution with time-varying parameter $\lambda_i(t/T)$: $X_{it} \sim P_{\lambda_i(t/T)}$.

Since $\lambda_i(t/T) = \mathbb{E}[X_{it}] = \mathrm{Var}(X_{it})$, we can rewrite X_{it} as

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + u_{it}\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it},$$

where η_{it} has zero mean and unit variance.

In applications the variance can be larger than the mean $\Rightarrow$ quasi-Poisson models.

Model

Quasi-Poisson model:

$$X_{it} = \lambda_i\Big(\frac{t}{T}\Big) + \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it},$$

where

- λ_i are unknown trend functions on $[0,1]$;
- σ is the overdispersion parameter;
- η_{it} are error terms that are independent across i and t and have zero mean and unit variance.

Testing procedure

Testing problem

Let $\mathcal{F} := \{\mathcal{I}_k \subseteq [0,1] : 1 \leq k \leq K\}$ be a family of rescaled time intervals on $[0,1]$, and for each triplet (i,j,k) consider the null hypothesis that the functions λ_i and λ_j are equal on an interval $\mathcal{I}_k$, i.e.

$$H_0^{(ijk)} : \quad \lambda_i(w) = \lambda_j(w) \text{ for all } w \in \mathcal{I}_k$$

We want to test $H_0^{(ijk)}$ simultaneously for all pairs of countries i and j and all intervals $\mathcal{I}_k$ in the family $\mathcal{F}$ and we want to control the familywise error rate (FWER) at level α:

$$\text{FWER}(\alpha) = \mathrm{P}\Big(\exists(i,j,k) : \text{we wrongly reject } H_0^{(ijk)}\Big)$$

Test statistic

For a given interval $\mathcal{I}_k$ and a pair of time series i and j we calculate

$$\hat{s}_{ijk} = \frac{1}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(X_{it} - X_{jt}),$$

where h_k is the length of $\mathcal{I}_k$. $\hat{s}_{ijk}$ estimates the average distance between λ_i and λ_j on $\mathcal{I}_k$:

$$\begin{aligned}\hat{s}_{ijk} =& \frac{1}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\bigg(\lambda_i\Big(\frac{t}{T}\Big) - \lambda_j\Big(\frac{t}{T}\Big)\bigg) \\ &+ \frac{\sigma}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\bigg(\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it} - \sqrt{\lambda_j\Big(\frac{t}{T}\Big)}\eta_{jt}\bigg) \\ =& \frac{1}{Th_k} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\bigg(\lambda_i\Big(\frac{t}{T}\Big) - \lambda_j\Big(\frac{t}{T}\Big)\bigg) + o_P(1)\end{aligned}$$

Test statistic, part 2

Under certain assumptions,

$$\mathrm{Var}(\hat{s}_{ijk}) = \frac{\sigma^2}{T^2 h_k^2} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big) \Big\{ \lambda_i\Big(\frac{t}{T}\Big) + \lambda_j\Big(\frac{t}{T}\Big) \Big\}$$

In order to normalize the variance of the statistic $\hat{s}_{ijk}$, we scale it by an estimator of its variance:

$$\widehat{\mathrm{Var}(\hat{s}_{ijk})} = \frac{\hat{\sigma}^2}{T^2 h_k^2} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big) (X_{it} + X_{jt}),$$

with $\hat{\sigma}^2$ being an appropriate estimator of σ^2. Details

Test statistic, part 3

Test statistic for the hypothesis $H_0^{(ijk)}$ is defined as

$$\widehat{\psi}_{ijk} := \frac{\hat{s}_{ijk}}{\sqrt{\widehat{\mathrm{Var}}(\hat{s}_{ijk})}} = \frac{\sum_{t=1}^{T} 1\left(\frac{t}{T} \in \mathcal{I}_k\right)(X_{it} - X_{jt})}{\hat{\sigma}\left\{\sum_{t=1}^{T} 1\left(\frac{t}{T} \in \mathcal{I}_k\right)(X_{it} + X_{jt})\right\}^{1/2}}$$

Critical values

How to construct critical values $c_{ijk}(\alpha)$?

- Traditional approach: $c_{ijk}(\alpha) = c(\alpha)$ for all (i, j, k).
- More modern approach: $c_{ijk}(\alpha)$ depend on the length h_k of the time interval (Dümbgen and Spokoiny (2001)):

$$c_{ijk}(\alpha) = c(\alpha, h_k) := b_k + q(\alpha)/a_k,$$

where a_k and b_k are scale-dependent constants and $q(\alpha)$ is chosen such that we control FWER. Details

Critical values, part 2

We want to control FWER. Let $\mathcal{M}_0 := \big\{(i,j,k) | H_0^{(ijk)} \text{ is true}\big\}$, then

$$\begin{aligned}
\text{FWER}(\alpha) &= \mathrm{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\widehat{\psi}_{ijk}| > c_{ijk}(\alpha)\Big) \\
&= 1 - \mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\widehat{\psi}_{ijk}| \le c_{ijk}(\alpha)\Big) \\
&= 1 - \mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : a_k\big(|\hat{\psi}_{ijk}| - b_k\big) \le q(\alpha)\Big) \\
&= 1 - \mathrm{P}\Big(\max_{(i,j,k) \in \mathcal{M}_0} a_k\big(|\hat{\psi}_{ijk}| - b_k\big) \le q(\alpha)\Big) \\
&\le 1 - \mathrm{P}\Big(\max_{(i,j,k)} a_k\big(|\hat{\psi}^0_{ijk}| - b_k\big) \le q(\alpha)\Big)
\end{aligned}$$

Hence, we choose $q(\alpha)$ as the $(1-\alpha)$-quantile of the statistic

$$\hat{\Psi}_T = \max_{(i,j,k)} a_k\big(|\hat{\psi}^0_{ijk}| - b_k\big),$$

where $\hat{\psi}^0_{ijk}$ is equal to $\hat{\psi}_{ijk}$ under the null.

Critical values, part 3

But we do not know the distribution of $\hat{\Psi}_T$ in practice!

$\Rightarrow$ the quantiles $q(\alpha)$ are also not known. How to approximate them?

Under our assumptions,

$$\hat{\psi}^0_{ijk} \approx \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt}),$$

which can be approximated by a Gaussian version of the test statistic:

$$\phi_{ijk} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(Z_{it} - Z_{jt}),$$

where Z_{it} are independent standard normal random variables.

Test procedure

1. Consider the Gaussian test statistic

$$\Phi_T = \max_{(i,j,k)} a_k\big(|\phi_{ijk}| - b_k\big),$$

   where a_k and b_k are scale-dependent constants and ϕ_{ijk} are weighted averages of the differences of standard normal random variables.

2. Compute a $(1-\alpha)$-quantile $q_{\text{Gauss}}(\alpha)$ of Φ_T by Monte Carlo simulations.

3. Adjust $q_{\text{Gauss}}(\alpha)$ by the scale-dependent constants

$$c_{\text{Gauss}}(\alpha, h_k) = b_k + q_{\text{Gauss}}(\alpha)/a_k$$

Test procedure

For the given significance level $\alpha \in (0,1)$ and for each (i,j,k), reject $H_0^{(ijk)}$ if $|\widehat{\psi}_{ijk}| > c_{\text{Gauss}}(\alpha, h_k)$.

Theoretical properties

Assumptions

$\mathcal{C}1$ The functions λ_i are uniformly Lipschitz continuous:
$|\lambda_i(u) - \lambda_i(v)| \leq L|u - v|$ for all $u, v \in [0, 1]$.

$\mathcal{C}2$ $0 < \lambda_{\min} \leq \lambda_i(w) \leq \lambda_{\max} < \infty$ for all $w \in [0, 1]$ and all i.

$\mathcal{C}3$ η_{it} are independent both across i and t.

$\mathcal{C}4$ $\mathbb{E}[\eta_{it}] = 0$, $\mathbb{E}[\eta_{it}^2] = 1$ and $\mathbb{E}[|\eta_{it}|^\theta] \leq C_\theta < \infty$ for some $\theta > 4$.

$\mathcal{C}5$ $h_{\max} = o(1/\log T)$ and $h_{\min} \geq CT^{-b}$ for some $b \in (0, 1)$.

$\mathcal{C}6$ $p := \{\#(i, j, k)\} = O(T^{(\theta/2)(1-b)-(1+\delta)})$ for some small $\delta > 0$.

Theoretical properties

Proposition

Let $\mathcal{M}_0$ be the set of triplets (i,j,k) for which $H_0^{(ijk)}$ holds true. Then under $\mathcal{C}1-\mathcal{C}6$, it holds that

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk}| \le c_{\text{Gauss}}(\alpha, h_k)\Big) \ge 1 - \alpha + o(1)$$

Proposition

Consider a sequence of functions $\lambda_i = \lambda_{i,T}$, $\lambda_j = \lambda_{j,T}$ such that

$$\exists \mathcal{I}_k : \lambda_i(w) - \lambda_j(w) \ge c_T \sqrt{\log T/(Th_k)} \; \forall w \in \mathcal{I}_k, \tag{1}$$

and $c_T \to \infty$ faster than $\frac{\sqrt{\log T}\sqrt{\log \log T}}{\log \log \log T}$. Let $\mathcal{M}_1$ be the set of triplets (i,j,k) for which (1) holds true. Then under $\mathcal{C}1-\mathcal{C}6$, it holds that

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_1 : |\hat{\psi}_{ijk}| > c_{\text{Gauss}}(\alpha, h_k)\Big) = 1 - o(1)$$

Application

Graphical representation

How to represent the results of the test?

Plot the results of pairwise comparison $\mathcal{F}_{\text{reject}}(i,j)$:

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : \mathcal{I}_k \notin \mathcal{F}_{\text{reject}}(i,j)\Big) \geq 1 - \alpha + o(1)$$

Minimal intervals

An interval $\mathcal{I}_k \in \mathcal{F}_{\text{reject}}(i,j)$ is called **minimal** if there is no other interval $\mathcal{I}_{k'} \in \mathcal{F}_{\text{reject}}(i,j)$ with $\mathcal{I}_{k'} \subset \mathcal{I}_k$.

The set of minimal intervals is denoted $\mathcal{F}^{\min}_{\text{reject}}(i,j)$.

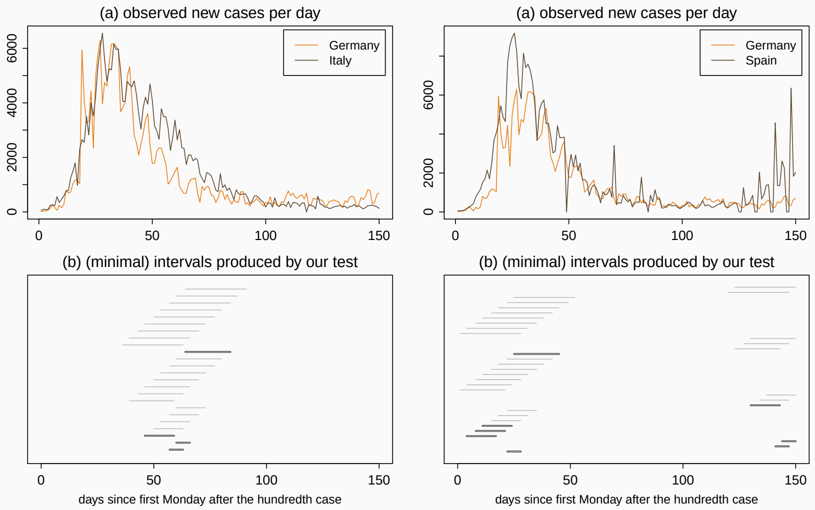
We can make similar confidence statements about minimal intervals:

$$\mathrm{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : \mathcal{I}_k \notin \mathcal{F}^{\min}_{\text{reject}}(i,j)\Big) \geq 1 - \alpha + o(1)$$

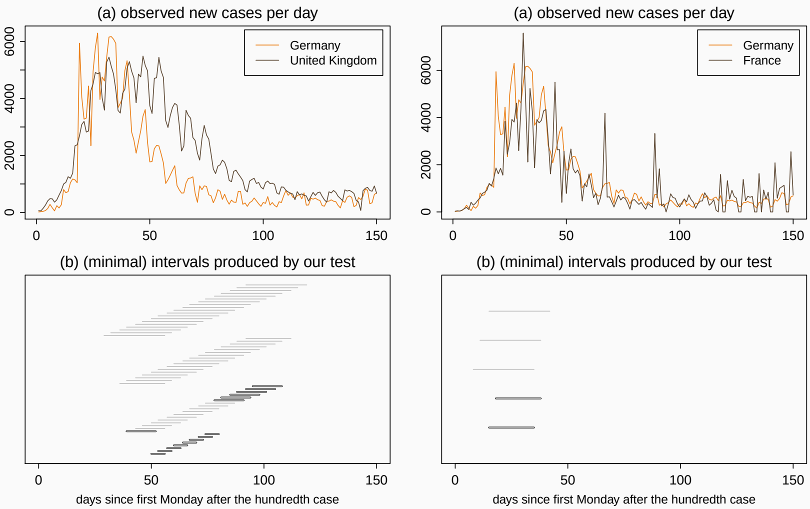
Application setting

- Five countries: Germany, Italy, Spain, France and the UK.
- $T = 150$ days.
- The data is aligned by weekdays: first Monday after reaching 100 cases as $t = 1$.
- Lengths of time intervals $7, 14, 21, 28$ days. The intervals start at days $1, 8, 15, \ldots$ and $4, 11, 19, \ldots$
- $\alpha = 0.05$.
- 5000 Monte Carlo simulation runs to produce critical values.

Application results



Application results, part 2



Discussion

We can claim, with confidence of about 95%, that the null hypothesis is violated for all intervals (and all pairs of countries) for which our test rejects the null.

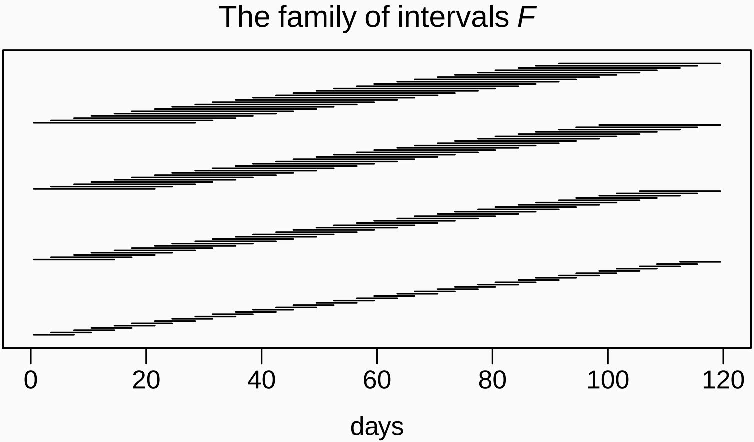
However, we can not say anything about the causes of such differences. This question requires further (probably not purely statistical) analysis.

Further possible extensions:

- introduce scaling factor in the trend function, that will allow to adjust for the size of the country (population, density, testing regimes, etc.);
- include dependence in the error terms;
- cluster the countries based on the trends they exhibit.

Thank you!

Family of time intervals



Simulation results for the size of the test

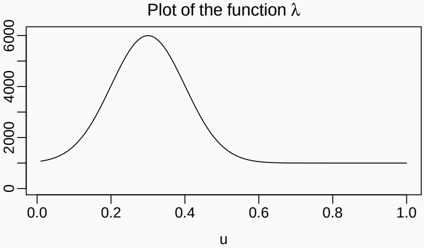
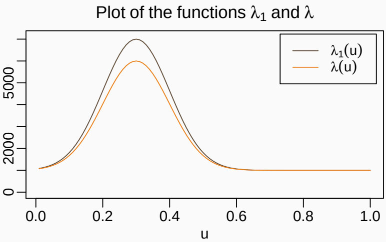


Table 1: Size of the multiscale test

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	0.011	0.047	0.093	0.010	0.044	0.087	0.008	0.037	0.075
$T = 250$	0.009	0.047	0.091	0.009	0.046	0.087	0.008	0.035	0.069
$T = 500$	0.010	0.044	0.083	0.008	0.048	0.093	0.007	0.035	0.077

Simulation results for the power of the test



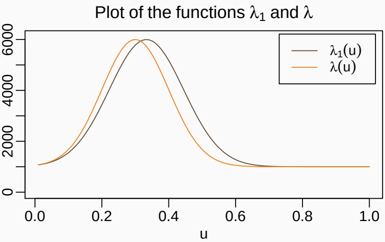


Table 3: Power of the multiscale test for scenario B

	$n = 5$			$n = 10$			$n = 50$		
	significance level α			significance level α			significance level α		
	0.01	0.05	0.1	0.01	0.05	0.1	0.01	0.05	0.1
$T = 100$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
$T = 250$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
$T = 500$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Estimator of σ^2

We estimate the overdispersion paramter σ^2 by

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_i^2 \text{ and } \hat{\sigma}_i^2 = \frac{\sum_{t=2}^{T}(X_{it} - X_{it-1})^2}{2\sum_{t=1}^{T} X_{it}}$$

We assume that λ_i is Lipschitz continuous. Then

$$X_{it} - X_{it-1} = \sigma\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}(\eta_{it} - \eta_{it-1}) + r_{it},$$

where $|r_{it}| \leq C(1 + |\eta_{it-1}|)/T$ with a sufficiently large C. Hence,

$$\frac{1}{T}\sum_{t=2}^{T}(X_{it} - X_{it-1})^2 = 2\sigma^2\Big\{\frac{1}{T}\sum_{t=2}^{T}\lambda_i(t/T)\Big\} + o_p(1)$$

Together with

$$\frac{1}{T}\sum_{t=1}^{T} X_{it} = \frac{1}{T}\sum_{t=1}^{T}\lambda_i(t/T) + o_p(1),$$

we get that $\hat{\sigma}_i^2 = \sigma^2 + o_p(1)$ for any i and thus $\hat{\sigma}^2 = \sigma^2 + o_p(1)$. Go back

Notation

In order to proceed with the proof, we will need the following notation:

$$\hat{\psi}_{ijk,T} = \frac{\sum_{t=1}^{T} 1\big(\frac{t}{T} \in \mathcal{I}_k\big)(X_{it} - X_{jt})}{\hat{\sigma}\big\{\sum_{t=1}^{T} 1\big(\frac{t}{T} \in \mathcal{I}_k\big)(X_{it} + X_{jt})\big\}^{1/2}}$$

$$\hat{\psi}^0_{ijk,T} = \frac{\sum_{t=1}^{T} 1\big(\frac{t}{T} \in \mathcal{I}_k\big)\,\sigma \overline{\lambda}_{ij}^{-1/2}(\frac{t}{T})(\eta_{it} - \eta_{jt})}{\hat{\sigma}\{\sum_{t=1}^{T} 1\big(\frac{t}{T} \in \mathcal{I}_k\big)(X_{it} + X_{jt})\}^{1/2}} \qquad \hat{\Psi}^0_T = \max_{(i,j,k)} a_k\big(|\hat{\psi}^0_{ijk,T}| - b_k\big)$$

$$\psi^0_{ijk,T} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(\eta_{it} - \eta_{jt}) \qquad \Psi_T = \max_{(i,j,k)} a_k\big(|\psi^0_{ijk,T}| - b_k\big)$$

$$\phi_{ijk,T} = \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(Z_{it} - Z_{jt}) \qquad \Phi_T = \max_{(i,j,k)} a_k\big(|\phi_{ijk,T}| - b_k\big)$$

Strategy of the proof

1. We prove that $\left|\hat{\Psi}_T^0 - \Psi_T\right| = o_p(r_T)$, where $\{r_T\}$ is some null sequence.
2. With the help of results from Chernozhukov et al. (2017), we prove

$$\sup_{q \in \mathrm{R}} \left| \mathrm{P}\big(\Psi_T \le q\big) - \mathrm{P}\big(\Phi_T \le q\big) \right| = o(1)$$

3. By using these two results, we now show that

$$\sup_{q \in \mathbb{R}} \left| \mathrm{P}\big(\hat{\Psi}_T^0 \le q\big) - \mathrm{P}\big(\Phi_T \le q\big) \right| = o(1) \tag{2}$$

4. It can be shown that $\mathrm{P}(\Phi_T \le q_{\mathsf{Gauss}}(\alpha)) = 1 - \alpha$. From this and (2), it immediately follows that

$$\mathrm{P}\big(\hat{\Psi}_T^0 \le q_{\mathsf{Gauss}}(\alpha)\big) = 1 - \alpha + o(1),$$

which in turn implies the desired statement.

Idea behind a_k and b_k

Dümbgen and Spokoiny (2001): the critical values $c_{ijk}(\alpha)$ depend on the scale of the testing problem, i.e. the length h_k of the time interval.

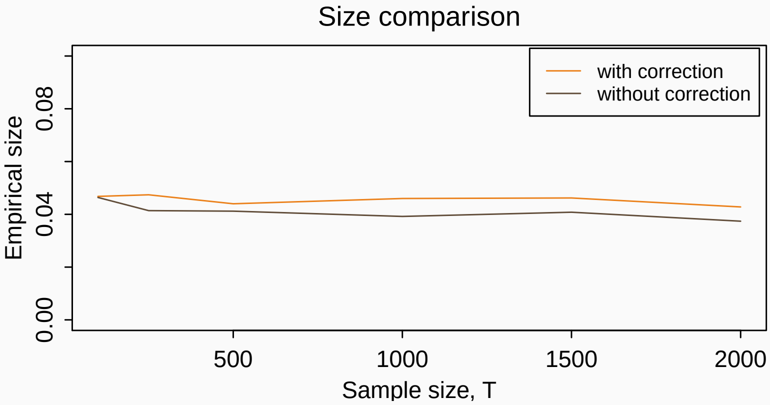
Specifically,

$$c_{ijk}(\alpha) = c(\alpha, h_k) := b_k + q(\alpha)/a_k,$$

where $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants and $q(\alpha)$ is chosen such that we control FWER.

Idea behind a_k and b_k, part 2

This choice of scale-dependent constants helps us balance the significance of hypotheses between the time intervals of different lengths h_k:



Go back

Idea behind the additive correction

Consider the uncorrected Gaussian statistic

$$\Phi^{\text{uncor}} = \max_{(i,j,k)} |\phi_{ijk}|$$

and let the family of intervals be

$$\mathcal{F} = \Big\{ [(m-1)h_l, mh_l] \text{ for } 1 \leq m \leq 1/h_l, 1 \leq l \leq L \Big\}$$

Then we can rewrite the uncorrected test statistic as

$$\Phi^{\text{uncor}} = \max_{i,j} \max_{\substack{1 \leq l \leq L, \\ 1 \leq m \leq 1/h_l}} \left| \frac{1}{\sqrt{2Th_l}} \sum_{t=1}^{T} 1\Big(\frac{t}{T} \in [(m-1)h_l, mh_l]\Big)(Z_{it} - Z_{jt}) \right|$$

$\Rightarrow \quad \max_m \ldots = \sqrt{2\log(1/h_l)} + o_P(1) \to \infty$ as $h \to 0$ and the stochastic behavior of Φ^{uncor} is dominated by the elements with small bandwidths h_l. Go back