Multiscale testing for equality of nonparametric trend curves

Marina Khismatullina
Erasmus University Rotterdam

Michael Vogt
University of Ulm

June 28, 2024

ISNPS 2024, Braga

Introduction

Motivation

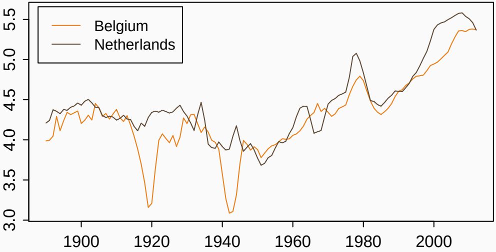
Aim of the paper

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

Motivation

Aim of the paper

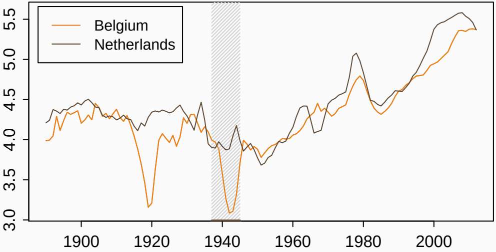
To develop new inference methods that allow to *identify* and *locate* differences between nonparametric trend curves with dependent errors.



Motivation

Aim of the paper

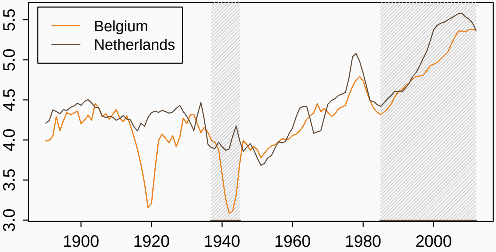
To develop new inference methods that allow to *identify* and *locate* differences between nonparametric trend curves with dependent errors.



Motivation

Aim of the paper

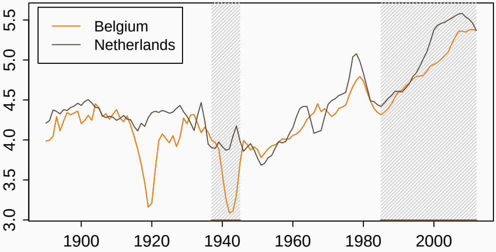
To develop new inference methods that allow to *identify* and *locate* differences between nonparametric trend curves with dependent errors.



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.

Motivation

Why is it relevant?

Finding systematic differences between trends = basis for further research.

Why is it difficult?

Testing many hypotheses at the same time = multiple testing problem

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

Motivation

Why is it relevant?

Finding systematic differences between trends = basis for further research.

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 a particular application?

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).

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).

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), Khismatullina and Vogt (2020).

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), Khismatullina and Vogt (2020).

Model

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$.

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

We consider the following model:

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

We consider the following model:

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

where

- m_i are unknown trend functions on $[0, 1]$;

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

We consider the following model:

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

where

- m_i are unknown trend functions on $[0, 1]$;
- $\boldsymbol{\beta}_i$ is $d \times 1$ vector of unknown parameters;

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

We consider the following model:

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

where

- m_i are unknown trend functions on $[0, 1]$;
- $\boldsymbol{\beta}_i$ is $d \times 1$ vector of unknown parameters;
- α_i are so-called fixed effect error terms;

Model

We observe a panel of n time series $\mathcal{Z}_i = \{(Y_{it}, \boldsymbol{X}_{it}) : 1 \leq t \leq T\}$ of length T, where $Y_{it} \in \mathbb{R}$ and $\boldsymbol{X}_{it} \in \mathbb{R}^d$. We assume that n is fixed.

We consider the following model:

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

where

- m_i are unknown trend functions on $[0, 1]$;
- $\boldsymbol{\beta}_i$ is $d \times 1$ vector of unknown parameters;
- α_i are so-called fixed effect error terms;
- $\mathcal{E}_i = \{\varepsilon_{it} : 1 \leq t \leq T\}$ is a zero-mean stationary and causal error process.

Model, part 2

$$Y_{it} = m_i\left(\frac{t}{T}\right) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

Model, part 2

$$Y_{it} = m_i\left(\frac{t}{T}\right) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

If we knew α_i and $\boldsymbol{\beta}_i$, then the model becomes much simpler:

$$\begin{aligned} Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} &=: Y_{it}^\circ \\ &= m_i\left(\frac{t}{T}\right) + \varepsilon_{it}. \end{aligned}$$

Model, part 2

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

If we knew α_i and $\boldsymbol{\beta}_i$, then the model becomes much simpler:

$$\begin{aligned} Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} &=: Y_{it}^\circ \\ &= m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}. \end{aligned}$$

In reality the variables Y_{it}° are **not** observed.

Model, part 2

$$Y_{it} = m_i\Big(\frac{t}{T}\Big) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

If we knew α_i and $\boldsymbol{\beta}_i$, then the model becomes much simpler:

$$\begin{aligned} Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} &=: Y_{it}^\circ \\ &= m_i\Big(\frac{t}{T}\Big) + \varepsilon_{it}. \end{aligned}$$

In reality the variables Y_{it}° are **not** observed.

But given $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$, we consider an augmented time series

$$\widehat{Y}_{it} := Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \boldsymbol{X}_{it} + m_i\Big(\frac{t}{T}\Big) + \big(\alpha_i - \widehat{\alpha}_i\big) + \varepsilon_{it}.$$

Model, part 2

$$Y_{it} = m_i\left(\frac{t}{T}\right) + \boldsymbol{\beta}_i^T \boldsymbol{X}_{it} + \alpha_i + \varepsilon_{it},$$

If we knew α_i and $\boldsymbol{\beta}_i$, then the model becomes much simpler:

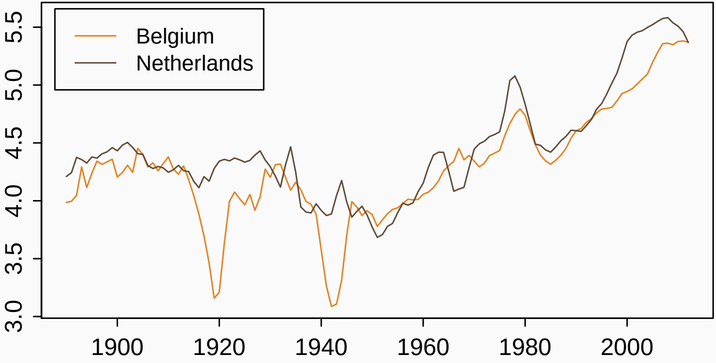
$$\begin{aligned} Y_{it} - \alpha_i - \boldsymbol{\beta}_i^\top \boldsymbol{X}_{it} &=: Y_{it}^\circ \\ &= m_i\left(\frac{t}{T}\right) + \varepsilon_{it}. \end{aligned}$$

In reality the variables Y_{it}° are **not** observed.

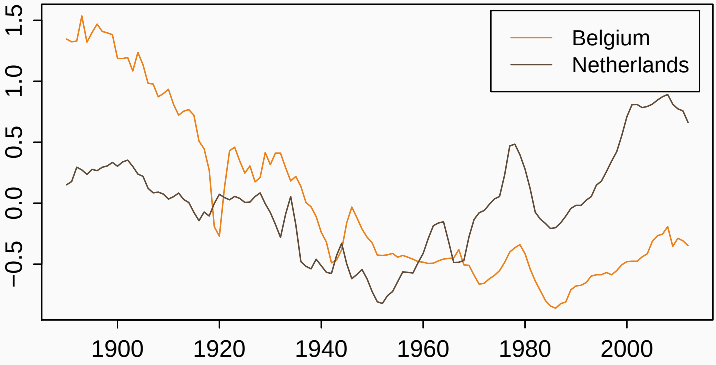
But given $\widehat{\alpha}_i$ and $\widehat{\boldsymbol{\beta}}_i$, we consider an augmented time series

$$\widehat{Y}_{it} := Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} = (\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i)^\top \boldsymbol{X}_{it} + m_i\left(\frac{t}{T}\right) + \left(\alpha_i - \widehat{\alpha}_i\right) + \varepsilon_{it}.$$

Original time series: Belgium and Netherlands



Augmented time series: Belgium and Netherlands



Testing procedure

Testing problem

$$H_0 : m_1 = m_2 = \ldots = m_n$$

Testing problem

$$H_0 : m_1 = m_2 = \ldots = m_n$$

Question: if we reject the global null, how to locate the differences between the trends?

Testing problem

$$H_0 : m_1 = m_2 = \ldots = m_n$$

Question: if we reject the global null, how to locate the differences between the trends?

Consider a grid $\mathcal{G}_T = \{(u, h) : [u - h, u + h] \subseteq [0, 1]\}$ of location-bandwidth parameters.

Testing problem

$$H_0 : m_1 = m_2 = \ldots = m_n$$

Question: if we reject the global null, how to locate the differences between the trends?

Consider a grid $\mathcal{G}_T = \{(u, h) : [u - h, u + h] \subseteq [0, 1]\}$ of location-bandwidth parameters. For each pair (i, j) and for each interval $[u - h, u + h]$ we consider the null hypothesis

$$H_0^{[i,j]}(u, h) : m_i(w) = m_j(w) \text{ for all } w \in [u - h, u + h].$$

Testing problem

$$H_0 : m_1 = m_2 = \ldots = m_n$$

Question: if we reject the global null, how to locate the differences between the trends?

Consider a grid $\mathcal{G}_T = \{(u, h) : [u - h, u + h] \subseteq [0, 1]\}$ of location-bandwidth parameters. For each pair (i, j) and for each interval $[u - h, u + h]$ we consider the null hypothesis

$$H_0^{[i,j]}(u, h) : m_i(w) = m_j(w) \text{ for all } w \in [u - h, u + h].$$

Then the global null $H_0 : m_1 = m_2 = \ldots = m_n$ can be reformulated as

$$H_0 : \text{The hypotheses } H_0^{[i,j]}(u, h) \text{ hold true for all intervals}$$
$$[u - h, u + h], (u, h) \in \mathcal{G}_T, \text{ and for all } 1 \leq i < j \leq n.$$

Test statistic

For a given location $u \in [0,1]$ and bandwidth h and a given pair (i,j) we construct the kernel averages

$$\widehat{\psi}_{ij}(u,h) = \sum_{t=1}^{T} w_t(u,h)\big(\widehat{Y}_{it} - \widehat{Y}_{jt}\big),$$

where $w_t(u,h)$ are appropriate weights.

Test statistic

For a given location $u \in [0,1]$ and bandwidth h and a given pair (i,j) we construct the kernel averages

$$\widehat{\psi}_{ij}(u,h) = \sum_{t=1}^{T} w_t(u,h)\big(\widehat{Y}_{it} - \widehat{Y}_{jt}\big),$$

where $w_t(u,h)$ are appropriate weights.

The kernel averages $\widehat{\psi}_{ij}(u,h)$ measure the distance between two trend curves m_i and m_j on $[u-h, u+h]$.

Test statistic, part 2

Instead with working directly with $\widehat{\psi}_{ij}(u,h)$, we replace them by

$$\widehat{\psi}^0_{ij}(u,h) = \left\{ \left| \frac{\widehat{\psi}_{ij}(u,h)}{\left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

Test statistic, part 2

Instead with working directly with $\widehat{\psi}_{ij}(u, h)$, we replace them by

$$\widehat{\psi}^0_{ij}(u, h) = \left\{ \left| \frac{\widehat{\psi}_{ij}(u, h)}{\left(\widehat{\sigma}^2_i + \widehat{\sigma}^2_j\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\widehat{\sigma}^2_i$ is an appropriate estimator of the long-run variance σ^2_i;

Test statistic, part 2

Instead with working directly with $\widehat{\psi}_{ij}(u,h)$, we replace them by

$$\widehat{\psi}^0_{ij}(u,h) = \left\{ \left| \frac{\widehat{\psi}_{ij}(u,h)}{\left(\widehat{\sigma}_i^2 + \widehat{\sigma}_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\widehat{\sigma}_i^2$ is an appropriate estimator of the long-run variance σ_i^2;
- $\lambda(h) = \sqrt{2\log\{1/(2h)\}}$ is an additive correction term (Dümbgen and Spokoiny (2001)).

Test statistic, part 3

To test the global null, we aggregate the individual test statistics for all (i, j) and all location-bandwidth pairs $(u, h) \in \mathcal{G}_T$:

$$\widehat{\Psi}_T = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\psi}^0_{ij}(u, h).$$

Test statistic, part 3

To test the global null, we aggregate the individual test statistics for all (i, j) and all location-bandwidth pairs $(u, h) \in \mathcal{G}_T$:

$$\widehat{\Psi}_T = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \widehat{\psi}^0_{ij}(u, h).$$

Main theoretical result

Under certain conditions and under the null, $\widehat{\psi}^0_{ij}(u, h)$ and $\widehat{\Psi}_T$ can be approximated by the corresponding Gaussian versions of the test statistics.

Test procedure

Gaussian version of the individual test statistics:

$$\phi_{ij}^0(u,h) = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_{ij}(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

Test procedure

Gaussian version of the individual test statistics:

$$\phi_{ij}^0(u,h) = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_{ij}(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\phi_{ij}(u,h) = \sum_{t=1}^{T} w_t(u,h) \left\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \right\};$

Test procedure

Gaussian version of the individual test statistics:

$$\phi_{ij}^0(u,h) = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_{ij}(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\phi_{ij}(u,h) = \sum_{t=1}^{T} w_t(u,h) \left\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \right\}$;
- Z_{it} are independent (across i and t) standard normal random variables;

Test procedure

Gaussian version of the individual test statistics:

$$\phi_{ij}^0(u,h) = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_{ij}(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\phi_{ij}(u,h) = \sum_{t=1}^{T} w_t(u,h) \left\{ \sigma_i(Z_{it} - \bar{Z}_i) - \sigma_j(Z_{jt} - \bar{Z}_j) \right\}$;
- Z_{it} are independent (across i and t) standard normal random variables;
- $\bar{Z}_i$ is the empirical average of $Z_{i1}, \dots, Z_{iT}$.

Test procedure

Gaussian version of the individual test statistics:

$$\phi_{ij}^0(u,h) = \max_{(u,h)\in\mathcal{G}_T} \left\{ \left| \frac{\phi_{ij}(u,h)}{\left(\sigma_i^2 + \sigma_j^2\right)^{1/2}} \right| - \lambda(h) \right\},$$

where

- $\phi_{ij}(u,h) = \sum_{t=1}^{T} w_t(u,h) \left\{ \sigma_i (Z_{it} - \bar{Z}_i) - \sigma_j (Z_{jt} - \bar{Z}_j) \right\}$;
- Z_{it} are independent (across i and t) standard normal random variables;
- $\bar{Z}_i$ is the empirical average of $Z_{i1}, \dots, Z_{iT}$.

Aggregated Gaussian test statistics:

$$\Phi_T = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \phi_{ij}^0(u,h).$$

Test procedure, part 2

1. Consider the Gaussian test statistic

$$\Phi_T = \max_{1 \leq i < j \leq n} \max_{(u,h) \in \mathcal{G}_T} \phi_{ij}^0(u, h),$$

   where ϕ_{ij}^0 are weighted averages of the differences of standard normal random variables.

Test procedure, part 2

1. Consider the Gaussian test statistic

$$\Phi_T = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \phi^0_{ij}(u,h),$$

   where ϕ^0_{ij} are weighted averages of the differences of standard normal random variables.
2. Compute a $(1-\alpha)$-quantile $q_T(\alpha)$ of Φ_T by Monte Carlo simulations.

Test procedure, part 2

1. Consider the Gaussian test statistic
$$\Phi_T = \max_{1\le i<j\le n} \max_{(u,h)\in\mathcal{G}_T} \phi_{ij}^0(u,h),$$
where ϕ_{ij}^0 are weighted averages of the differences of standard normal random variables.
2. Compute a $(1-\alpha)$-quantile $q_T(\alpha)$ of Φ_T by Monte Carlo simulations.
3. Perform the test for the global hypothesis H_0: reject H_0 if $\widehat{\Psi}_T > q_T(\alpha)$.

Test procedure, part 2

1. Consider the Gaussian test statistic

$$\Phi_T = \max_{1 \le i < j \le n} \max_{(u,h) \in \mathcal{G}_T} \phi_{ij}^0(u,h),$$

   where ϕ_{ij}^0 are weighted averages of the differences of standard normal random variables.

2. Compute a $(1-\alpha)$-quantile $q_T(\alpha)$ of Φ_T by Monte Carlo simulations.
3. Perform the test for the global hypothesis H_0: reject H_0 if $\widehat{\Psi}_T > q_T(\alpha)$.
4. For each i, j, and each $(u,h) \in \mathcal{G}_T$, carry out the test for the local null hypothesis $H_0^{[i,j]}(u,h)$: reject $H_0^{[i,j]}(u,h)$ if $\widehat{\psi}_{ij}^0(u,h) > q_T(\alpha)$.

Theoretical properties

Proposition

Under certain assumptions and under the null, it holds that

$$\mathbb{P}\Big(\widehat{\Psi}_T \leq q_T(\alpha)\Big) = 1 - \alpha + o(1)$$

Theoretical properties

Proposition

Under certain assumptions and under the null, it holds that

$$\mathbb{P}\Big(\widehat{\Psi}_T \leq q_T(\alpha)\Big) = 1 - \alpha + o(1)$$

Corollary

$$FWER(\alpha) \leq \alpha$$

Theoretical properties

Proposition

Under certain assumptions and under the null, it holds that

$$\mathbb{P}\Big(\widehat{\Psi}_T \leq q_T(\alpha)\Big) = 1 - \alpha + o(1)$$

Corollary

$$\mathit{FWER}(\alpha) \leq \alpha$$

Proposition

Consider a sequence of functions $m_i = m_{i,T}$, $m_j = m_{j,T}$ *such that*

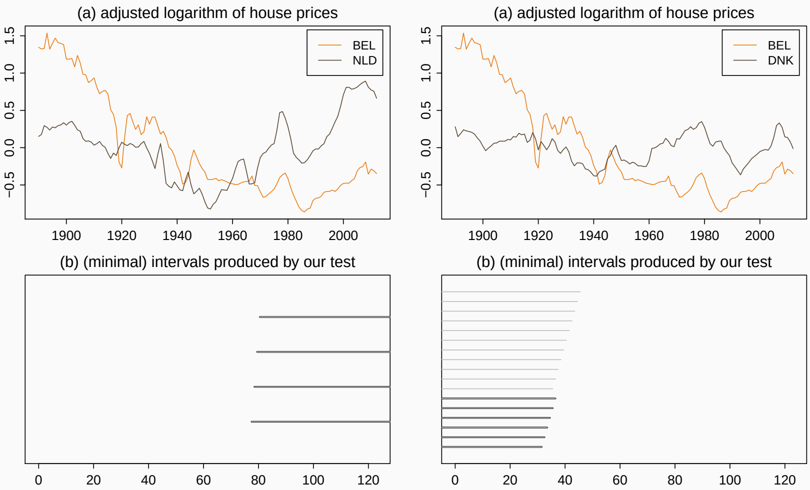
$$\exists\, (u,h) \in \mathcal{G}_T : m_i(w) - m_j(w) \geq c_T \sqrt{\log T/(Th)}\ \forall w \in [u-h, u+h],$$

and $c_T \to \infty$. *Then under our assumptions, it holds that*

$$\mathbb{P}\Big(\widehat{\Psi}_T \leq q_T(\alpha)\Big) = o(1)$$

Illustration

Application results



Discussion

We can claim, with confidence of at least 95%, that the null hypothesis is violated for all intervals (and all pairs of time series) 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:

- cluster the time series based on the trends they exhibit;
- introduce scaling factor in the trend function;
- include the dependence between covariates and error terms.

Thank you!

- R package: https://github.com/marina-khi/MSinference
- Khismatullina M., Vogt M. "Multiscale Comparison of Nonparametric Trend Curves", https://arxiv.org/abs/2209.10841

Model, part 3

1. We estimate $\boldsymbol{\beta}_i$:

$$\widehat{\boldsymbol{\beta}}_i = \Big(\sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta \boldsymbol{X}_{it}^\top \Big)^{-1} \sum_{t=2}^{T} \Delta \boldsymbol{X}_{it} \Delta Y_{it}$$

Theorem

Under certain regularity assumptions, $\widehat{\boldsymbol{\beta}}_i$ is a consistent estimator of $\boldsymbol{\beta}_i$ with the property $\boldsymbol{\beta}_i - \widehat{\boldsymbol{\beta}}_i = O_P(T^{-1/2})$.

2. We estimate the fixed effects α_i:

$$\widehat{\alpha}_i = \frac{1}{T} \sum_{t=1}^{T} \big(Y_{it} - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it} \big)$$

We then work with the augmented time series $\widehat{Y}_{it} = Y_{it} - \widehat{\alpha}_i - \widehat{\boldsymbol{\beta}}_i^\top \boldsymbol{X}_{it}$.

Test statistic

For a given location $u \in [0,1]$ and bandwidth h and a given pair (i,j) we construct the kernel averages

$$\widehat{\psi}_{ij,T}(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\big(\widehat{Y}_{it} - \widehat{Y}_{jt}\big),$$

where

$$w_{t,T}(u,h) = \frac{\Lambda_{t,T}(u,h)}{\{\sum_{t=1}^{T} \Lambda_{t,T}^2(u,h)\}^{1/2}},$$

$$\Lambda_{t,T}(u,h) = K\Big(\frac{t/T-u}{h}\Big)\Big[S_{T,2}(u,h) - S_{T,1}(u,h)\Big(\frac{t/T-u}{h}\Big)\Big],$$

$$S_{T,\ell}(u,h) = \frac{1}{Th}\sum_{t=1}^{T} K\Big(\frac{t/T-u}{h}\Big)\Big(\frac{t/T-u}{h}\Big)^{\ell}$$

for $\ell = 1, 2$ and K is a kernel function.

Assumptions

$\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

Assumptions

$\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

$\mathcal{C}2$ For each i the variables ε_{it} are weakly dependent. Details

Assumptions

- $\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.
- $\mathcal{C}2$ For each i the variables ε_{it} are weakly dependent. Details
- $\mathcal{C}3$ For each i we have that $\boldsymbol{X}_{it}$ is stationary and causal with all the necessary moments and no asymptotic multicollinearity.

Assumptions

$\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

$\mathcal{C}2$ For each i the variables ε_{it} are weakly dependent. Details

$\mathcal{C}3$ For each i we have that $\boldsymbol{X}_{it}$ is stationary and causal with all the necessary moments and no asymptotic multicollinearity.

$\mathcal{C}4$ For each i the variables $\boldsymbol{X}_{it}$ are weakly dependent. Details

Assumptions

$\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.

$\mathcal{C}2$ For each i the variables ε_{it} are weakly dependent. Details

$\mathcal{C}3$ For each i we have that $\boldsymbol{X}_{it}$ is stationary and causal with all the necessary moments and no asymptotic multicollinearity.

$\mathcal{C}4$ For each i the variables $\boldsymbol{X}_{it}$ are weakly dependent. Details

$\mathcal{C}5$ $\boldsymbol{X}_{it}$ (elementwise) and ε_{is} are uncorrelated for each t, s.

Assumptions

- $\mathcal{C}1$ For all i it holds that $\mathbb{E}[\varepsilon_{it}] = 0$ and $\|\varepsilon_{it}\|_q < \infty$ for some $q > 4$.
- $\mathcal{C}2$ For each i the variables ε_{it} are weakly dependent. Details
- $\mathcal{C}3$ For each i we have that $\boldsymbol{X}_{it}$ is stationary and causal with all the necessary moments and no asymptotic multicollinearity.
- $\mathcal{C}4$ For each i the variables $\boldsymbol{X}_{it}$ are weakly dependent. Details
- $\mathcal{C}5$ $\boldsymbol{X}_{it}$ (elementwise) and ε_{is} are uncorrelated for each t, s.
- $\mathcal{C}6$ All of the variables in the model are short-range dependent. Details

Assumptions, part 2

$\mathcal{C}7$ Standard assumptions on the kernel function K.

Assumptions, part 2

$\mathcal{C}7$ Standard assumptions on the kernel function K.

$\mathcal{C}8$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

Assumptions, part 2

$\mathcal{C}7$ Standard assumptions on the kernel function K.

$\mathcal{C}8$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

$$\mathcal{G}_T = \Big\{(u,h) : u = t/T \text{ for some } 1 \le t \le T \text{ and } h \in [h_{\min}, h_{\max}] \\ \text{with } h = t/T \text{ for some } 1 \le t \le T\Big\},$$

Assumptions, part 2

$\mathcal{C}7$ Standard assumptions on the kernel function K.

$\mathcal{C}8$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

$$\mathcal{G}_T = \Big\{(u,h) : u = t/T \text{ for some } 1 \leq t \leq T \text{ and } h \in [h_{\min}, h_{\max}] \\ \text{with } h = t/T \text{ for some } 1 \leq t \leq T\Big\},$$

$\mathcal{C}9$ $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$ and $h_{\max} < 1/2$.

Assumptions, part 2

$\mathcal{C}7$ Standard assumptions on the kernel function K.

$\mathcal{C}8$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

$$\mathcal{G}_T = \Big\{(u,h) : u = t/T \text{ for some } 1 \le t \le T \text{ and } h \in [h_{\min}, h_{\max}] \\ \text{with } h = t/T \text{ for some } 1 \le t \le T\Big\},$$

$\mathcal{C}9$ $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$ and $h_{\max} < 1/2$.

$\mathcal{C}10$ Assume that $\sigma_i^2 = \sigma_j^2$ for all i,j and $\widehat{\sigma}_i^2 = \sigma_i^2 + o_p(\rho_T)$ with $\rho_T = o(\sqrt{h_{\min}}/\log T)$.

Strategy of the proof

- Introduce $\widehat{\Phi}_{n,T}$ that is close in distribution to $\widehat{\Psi}_{n,T}$ under the null.

Strategy of the proof

- Introduce $\widehat{\Phi}_{n,T}$ that is close in distribution to $\widehat{\Psi}_{n,T}$ under the null.
- Using strong approximation theory for dependent processes as derived in Berkes et al. (2014), replace $\widehat{\Phi}_{n,T}$ by $\widetilde{\Phi}_{n,T}$ with the same distribution and the property that
$$\left|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\right| = o_p(\delta_T),$$
where δ_T goes to 0 as $T \to \infty$ sufficiently fast.

Strategy of the proof

- Introduce $\widehat{\Phi}_{n,T}$ that is close in distribution to $\widehat{\Psi}_{n,T}$ under the null.
- Using strong approximation theory for dependent processes as derived in Berkes et al. (2014), replace $\widehat{\Phi}_{n,T}$ by $\widetilde{\Phi}_{n,T}$ with the same distribution and the property that

$$\left|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\right| = o_p(\delta_T),$$

  where δ_T goes to 0 as $T \to \infty$ sufficiently fast.
- Using the anti-concentration results for Gaussian random vectors (Chernozhukov et al. 2015), prove that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$, i.e.

$$\sup_{x \in \mathbb{R}} \mathbb{P}\left(|\Phi_{n,T} - x| \leq \delta_T\right) = o(1).$$

Strategy of the proof

- Introduce $\widehat{\Phi}_{n,T}$ that is close in distribution to $\widehat{\Psi}_{n,T}$ under the null.
- Using strong approximation theory for dependent processes as derived in Berkes et al. (2014), replace $\widehat{\Phi}_{n,T}$ by $\widetilde{\Phi}_{n,T}$ with the same distribution and the property that

$$\left|\widetilde{\Phi}_{n,T} - \Phi_{n,T}\right| = o_p(\delta_T),$$

  where δ_T goes to 0 as $T \to \infty$ sufficiently fast.
- Using the anti-concentration results for Gaussian random vectors (Chernozhukov et al. 2015), prove that $\Phi_{n,T}$ does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$, i.e.

$$\sup_{x \in \mathbb{R}} \mathbb{P}\big(|\Phi_{n,T} - x| \leq \delta_T\big) = o(1).$$

- Show that

$$\sup_{x \in \mathbb{R}} \left|\mathbb{P}(\widetilde{\Phi}_{n,T} \leq x) - \mathbb{P}(\Phi_{n,T} \leq x)\right| = o(1).$$

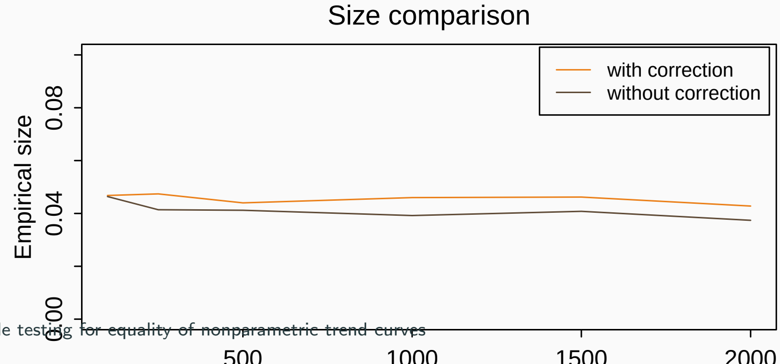
Idea behind $\lambda(h)$

Dümbgen and Spokoiny (2001): the critical values for testing the 'local' null hypothesis depend on the scale of the testing problem, i.e. the length h of the time interval.

Idea behind $\lambda(h)$

Dümbgen and Spokoiny (2001): the critical values for testing the 'local' null hypothesis depend on the scale of the testing problem, i.e. the length h of the time interval.

Introduction of a scale-dependent parameter helps us balance the significance of hypotheses between the time intervals of different lengths h_k:



Idea behind the additive correction

Consider the uncorrected Gaussian statistic

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

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

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

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

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

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

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

Then we can rewrite the uncorrected test statistic as

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

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

Dependence measure

Following Wu (2005), we define the *physical dependence measure* for the process $\boldsymbol{L}(\mathcal{F}_t)$ as the following:

$$\delta_q(\boldsymbol{L}, t) = ||\boldsymbol{L}(\mathcal{F}_t) - \boldsymbol{L}(\mathcal{F}_t')||_q,$$

where $\mathcal{F}_t = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ and $\mathcal{F}_t' = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ is a coupled process of $\mathcal{F}_t$ with ϵ_0' being an i.i.d. copy of ϵ_0.

Dependence measure

Following Wu (2005), we define the *physical dependence measure* for the process $\boldsymbol{L}(\mathcal{F}_t)$ as the following:

$$\delta_q(\boldsymbol{L}, t) = ||\boldsymbol{L}(\mathcal{F}_t) - \boldsymbol{L}(\mathcal{F}_t')||_q,$$

where $\mathcal{F}_t = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ and $\mathcal{F}_t' = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t)$ is a coupled process of $\mathcal{F}_t$ with ϵ_0' being an i.i.d. copy of ϵ_0.

Intuitively, $\delta_q(\boldsymbol{L}, t)$ measures the dependency of $\boldsymbol{L}(\mathcal{F}_t)$ on ϵ_0, i.e., how replacing ϵ_0 by an i.i.d. copy while keeping all other innovations in place affects the output $\boldsymbol{L}(\mathcal{F}_t)$.

Technical assumptions

$\mathcal{C}1'$ The variables ε_{it} are independent across i and allow for the representation $\varepsilon_{it} = G_i(\ldots, \eta_{it-1}, \eta_{it})$, where η_{it} are i.i.d. random variables across t and $G_i : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function..

$\mathcal{C}1''$ Define $\Theta_{i,t,q} = \sum_{s \geq t} \delta_q(G_i, s)$ for $t \geq 0$. For each i it holds that $\Theta_{i,t,q} = O(t^{-\tau_q}(\log t)^{-A})$, where $A > \frac{2}{3}(1/q + 1 + \tau_q)$ and $\tau_q = \{q^2 - 4 + (q-2)\sqrt{q^2 + 20q + 4}\}/8q$. Go back

Technical assumptions, part 2

$\mathcal{C}3'$ $\boldsymbol{X}_{it}$ allow for the representation $\boldsymbol{X}_{it} = \boldsymbol{H}_i(\ldots, u_{it-1}, u_{it})$ with u_{it} being i.i.d. random variables and $\boldsymbol{H}_i := (H_{i1}, H_{i2}, \ldots, H_{id})^\top : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^d$ being a measurable function such that $\boldsymbol{H}_i(\mathcal{U}_{it})$ is well defined.

$\mathcal{C}3''$ Let $\boldsymbol{N}_i$ be the $d \times d$ matrix with $n_{i,kl} = \mathbb{E}[H_{ik}(\mathcal{U}_{i0})H_{il}(\mathcal{U}_{i0})]$ being kl-th entry. We assume that the smallest eigenvalue of $\boldsymbol{N}_i$ is strictly bigger than 0.

$\mathcal{C}3'''$ Let $\mathbb{E}[\boldsymbol{H}_i(\mathcal{U}_{i0})] = \boldsymbol{0}$ and $||\boldsymbol{H}_i(\mathcal{U}_{it})||_{q'} < \infty$ for some $q' > \max\{2\theta, 4\}$, where θ will be introduced further.

$\mathcal{C}4'$ $\sum_{s=0}^{\infty} \delta_{q'}(\boldsymbol{H}_i, s) < \infty$ for q' from Assumption $\mathcal{C}3'''$.

$\mathcal{C}4''$ For each i it holds that $\sum_{s=t}^{\infty} \delta_{q'}(\boldsymbol{H}_i, s) = O(t^{-\alpha})$ for q' from Assumption $\mathcal{C}3'''$ and for some $\alpha > 1/2 - 1/q'$. Go back

Technical assumptions, part 3

$\mathcal{C}6$ Let $\zeta_{i,t} = (u_{it}, \eta_{it})^\top$. Denote $\mathcal{I}_{it} = (\ldots, \zeta_{i,t-1}, \zeta_{i,t})$, $\mathcal{J}_{it} = (\ldots, \eta_{it-2}, \eta_{it-1}, \eta_{it})$, $\mathcal{U}_{it} = (\ldots, u_{it-1}, u_{it})$, and $\mathbf{U}_i(\mathcal{I}_{it}) = \boldsymbol{H}_i(\mathcal{U}_{it}) G_i(\mathcal{J}_{it})$. With this notation at hand, we assume that $\sum_{s=0}^{\infty} \delta_2(\mathbf{U}_i, s) < \infty$. Go back

Graphical representation

How to represent the results of the test?

Graphical representation

How to represent the results of the test?

We can plot all of the intervals where we reject the local null.

Graphical representation

How to represent the results of the test?

We can plot all of the intervals where we reject the local null.

But what if there are too many?

Graphical representation

How to represent the results of the test?

We can plot all of the intervals where we reject the local null.

But what if there are too many?

An interval $[u-h, u+h]$ is called **minimal** if the corresponding local null $H_0^{[i,j]}(u,h)$ is rejected and there is no other interval $[u'-h', u'+h']$ such that we reject $H_0^{[i,j]}(u',h')$ and $[u'-h', u'+h'] \subset [u-h, u+h]$.