Simultaneous statistical inference for epidemic trends: the case of COVID-19

Marina Khismatullina Michael Vogt

01/10/2020

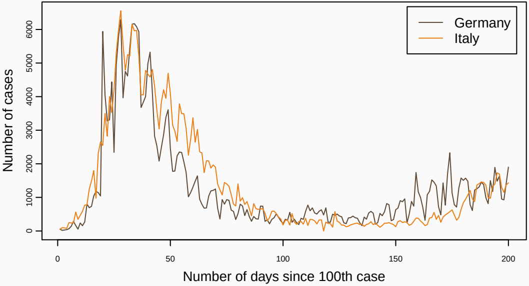
Table of contents

Introduction

Motivation

Research question:

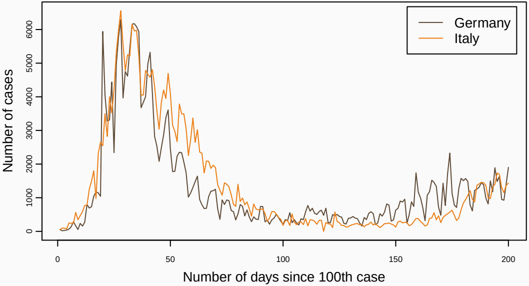
How do outbreak patterns of COVID-19 compare across countries?



Motivation

Research question:

How do outbreak patterns of COVID-19 compare across countries?



Aim of the paper

To develop new inference methods that allow to *identify* and *locate* differences between time trends.

Model

Model

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

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

Model

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

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

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.

Literature

Curve comparisons

- Park et al. (2009)

Literature

Curve comparisons

- Park et al. (2009)

Studies of COVID-19

- SEIR models

The multiscale testing method

Testing problem

Let $\mathcal{F} = \{\mathcal{I}_k \subseteq [0,1] : 1 \leq k \leq K\}$ be a family of intervals on $[0,1]$, and for a given interval $\mathcal{I}_k$ we want to test whether the functions λ_i and λ_j are the same on this interval. Formally, the testing problem is

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

Testing problem

Let $\mathcal{F} = \{\mathcal{I}_k \subseteq [0,1] : 1 \leq k \leq K\}$ be a family of intervals on $[0,1]$, and for a given interval $\mathcal{I}_k$ we want to test whether the functions λ_i and λ_j are the same on this interval. Formally, the testing problem is

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

We want to test these hypothesis $H_0^{(ijk)}$ simultaneously for all pairs of countries i and j and all intervals $\mathcal{I}_k$ in the family $\mathcal{F}$.

Test statistic

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

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

where h_k is the length of the interval $\mathcal{I}_k$.

Test statistic

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

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

where h_k is the length of the interval $\mathcal{I}_k$. Under certain assumptions,

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

Test statistic

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

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

where h_k is the length of the interval $\mathcal{I}_k$. Under certain assumptions,

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

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

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

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

Test statistic

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

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

Test statistic

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

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

Under certain conditions and under the null, $\widehat{\psi}_{ijk,T}$ can be approximated by the Gaussian version of the test statistic:

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

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

Test procedure

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

Test procedure

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

Test procedure

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

Test procedure

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

Test procedure

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

- $c_{T,\text{Gauss}}(\alpha, h_k) = b_k + q_{T,\text{Gauss}}(\alpha)/a_k$ is a scale-dependent critical value;

Test procedure

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

Test procedure

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

- $c_{T,\text{Gauss}}(\alpha, h_k) = b_k + q_{T,\text{Gauss}}(\alpha)/a_k$ is a scale-dependent critical value;
- $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants; Idea

Test procedure

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

Test procedure

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

- $c_{T,\text{Gauss}}(\alpha, h_k) = b_k + q_{T,\text{Gauss}}(\alpha)/a_k$ is a scale-dependent critical value;
- $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants; Idea
- $q_{T,\text{Gauss}}(\alpha)$ is $(1-\alpha)$-quantile of the Gaussian test statistic Φ_T;

Test procedure

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

Test procedure

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

- $c_{T,\text{Gauss}}(\alpha, h_k) = b_k + q_{T,\text{Gauss}}(\alpha)/a_k$ is a scale-dependent critical value;
- $a_k = \{\log(e/h_k)\}^{1/2}/\log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$ are scale-dependent constants; Idea
- $q_{T,\text{Gauss}}(\alpha)$ is $(1-\alpha)$-quantile of the Gaussian test statistic Φ_T;
- and $\Phi_T = \max_{(i,j,k)} a_k\big(|\phi_{ijk,T}| - b_k\big)$ the Gaussian test statistic.

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 $u, v \in [0, 1]$.

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.

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.

Assumptions

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

$\mathcal{C}2$ $0 < \lambda_{\text{min}} \le \lambda_i(w) \le \lambda_{\text{max}} < \infty$ 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 $\mathbb{E}[|\eta_{it}|^\theta] \le C_\theta < \infty$ for some $\theta > 4$.

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 $b \in (0, 1)$.

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

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

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

Strategy of the proof

- Replace the statistic $\widehat{\Psi}_T$ under $H_0 : m = 0$ by a statistic $\widetilde{\Phi}_T$ with the same distribution and the property that

$$\left|\widetilde{\Phi}_T - \Phi_T\right| = o_p(\delta_T),$$

where $\delta_T = o(1)$. To do so, we make use of strong approximation theory for dependent processes as derived in Berkes et al. (2014)

Strategy of the proof

- Replace the statistic $\widehat{\Psi}_T$ under $H_0 : m = 0$ by a statistic $\widetilde{\Phi}_T$ with the same distribution and the property that

$$\left|\widetilde{\Phi}_T - \Phi_T\right| = o_p(\delta_T),$$

  where $\delta_T = o(1)$. To do so, we make use of strong approximation theory for dependent processes as derived in Berkes et al. (2014)

- Using the anti-concentration results for Gaussian random vectors (Chernozhukov et al. 2015), prove that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$, i.e.

$$\sup_{x \in \mathbb{R}} \mathrm{P}\left(|\Phi_T - x| \leq \delta_T\right) = o(1).$$

Strategy of the proof

- Replace the statistic $\widehat{\Psi}_T$ under $H_0 : m = 0$ by a statistic $\widetilde{\Phi}_T$ with the same distribution and the property that

$$\left|\widetilde{\Phi}_T - \Phi_T\right| = o_p(\delta_T),$$

  where $\delta_T = o(1)$. To do so, we make use of strong approximation theory for dependent processes as derived in Berkes et al. (2014)

- Using the anti-concentration results for Gaussian random vectors (Chernozhukov et al. 2015), prove that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$, i.e.

$$\sup_{x \in \mathbb{R}} \mathrm{P}\Big(|\Phi_T - x| \leq \delta_T\Big) = o(1).$$

- Show that

$$\sup_{x \in \mathbb{R}} \left|\mathrm{P}(\widetilde{\Phi}_T \leq x) - \mathrm{P}(\Phi_T \leq x)\right| = o(1).$$

Theoretical properties

Define

$$\Pi_T^+ = \left\{ I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^+ \text{ and } I_{u,h} \subseteq [0,1] \right\}$$

with

$$\mathcal{A}_T^+ = \left\{ (u,h) \in \mathcal{G}_T : \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} > q_T(\alpha) + \lambda(h) \right\}$$

Theoretical properties

Define

$$\Pi_T^+ = \big\{ I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^+ \text{ and } I_{u,h} \subseteq [0,1] \big\}$$
$$\Pi_T^- = \big\{ I_{u,h} = [u-h, u+h] : (u,h) \in \mathcal{A}_T^- \text{ and } I_{u,h} \subseteq [0,1] \big\}$$

with

$$\mathcal{A}_T^+ = \Big\{ (u,h) \in \mathcal{G}_T : \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} > q_T(\alpha) + \lambda(h) \Big\}$$
$$\mathcal{A}_T^- = \Big\{ (u,h) \in \mathcal{G}_T : -\frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} > q_T(\alpha) + \lambda(h) \Big\}$$

Theoretical properties

Proposition

Under our assumptions, for events $E_T^+ = \left\{ \forall I_{u,h} \in \Pi_T^+ : m'(v) > 0 \text{ for some } v \in I_{u,h} \right\}$ *it holds that*

$$\mathrm{P}\big(E_T^+\big) \geq (1 - \alpha) + o(1)$$

Theoretical properties

Proposition

Under our assumptions, for events
$E_T^+ = \left\{\forall I_{u,h} \in \Pi_T^+ : m'(v) > 0 \textit{ for some } v \in I_{u,h}\right\}$ *and*
$E_T^- = \left\{\forall I_{u,h} \in \Pi_T^- : m'(v) < 0 \textit{ for some } v \in I_{u,h}\right\}$ *it holds that*

$$\mathrm{P}\big(E_T^+\big) \geq (1-\alpha) + o(1)$$
$$\mathrm{P}\big(E_T^-\big) \geq (1-\alpha) + o(1)$$

Graphical representation

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 $\mathcal{F}^{\text{min}}_{\text{reject}}(i,j)$.

Graphical representation

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 $\mathcal{F}^{\text{min}}_{\text{reject}}(i,j)$.

Define

$$\Pi_T^{min,+} = \text{ set of minimal intervals from } \Pi_T^+,$$
$$E_T^{min,+} = \Big\{ \forall I_{u,h} \in \Pi_T^{min,+} : m'(v) > 0 \text{ for some } v \in I_{u,h} \Big\}$$

Graphical representation

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 $\mathcal{F}_{\text{reject}}^{\text{min}}(i, j)$.

Define

$$\Pi_T^{min,+} = \text{ set of minimal intervals from } \Pi_T^+,$$
$$E_T^{min,+} = \left\{ \forall I_{u,h} \in \Pi_T^{min,+} : m'(v) > 0 \text{ for some } v \in I_{u,h} \right\}$$

Since $E_T^{min,+} = E_T^+$, we have

$$\mathrm{P}\left(E_T^{min,+}\right) \geq (1 - \alpha) + o(1).$$

Conclusion

Conclusion

We developed multiscale methods to test qualitative hypotheses about nonparametric time trends:

- whether the trend is present at all;
- whether the trend function is constant;
- in which time regions there is an upward/downward movement in the trend.

We derived asymptotic theory for the proposed tests.

As an application of our method, we analyzed the behavior of the yearly mean temperature in Central England from 1659 to 2017.

Thank you!

Long-run error variance estimator

Setting

Estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ of the error terms $\{\varepsilon_t\}$ in the model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a stationary and causal AR(p) process of the form

$$\varepsilon_t = \sum_{j=1}^{p} a_j \varepsilon_{t-j} + \eta_t.$$

Setting

Estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ of the error terms $\{\varepsilon_t\}$ in the model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a stationary and causal AR(p) process of the form

$$\varepsilon_t = \sum_{j=1}^{p} a_j \varepsilon_{t-j} + \eta_t.$$

- $\boldsymbol{a} = (a_1, \ldots, a_p)$ is a vector of the unknown parameters;

Setting

Estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ of the error terms $\{\varepsilon_t\}$ in the model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a stationary and causal AR(p) process of the form

$$\varepsilon_t = \sum_{j=1}^{p} a_j \varepsilon_{t-j} + \eta_t.$$

- $\boldsymbol{a} = (a_1, \ldots, a_p)$ is a vector of the unknown parameters;
- η_t are i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \nu^2$;

Setting

Estimate the long-run error variance $\sigma^2 = \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\varepsilon_0, \varepsilon_\ell)$ of the error terms $\{\varepsilon_t\}$ in the model

$$Y_t = m\Big(\frac{t}{T}\Big) + \varepsilon_t,$$

where $\{\varepsilon_t\}$ is a stationary and causal AR(p) process of the form

$$\varepsilon_t = \sum_{j=1}^{p} a_j \varepsilon_{t-j} + \eta_t.$$

- $\boldsymbol{a} = (a_1, \dots, a_p)$ is a vector of the unknown parameters;
- η_t are i.i.d. innovations with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \nu^2$;
- p is known.

Estimator, first stage

Yule-Walker equations yield

$$\boldsymbol{\Gamma}_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q,$$

where

- $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$;

Estimator, first stage

Yule-Walker equations yield

$$\boldsymbol{\Gamma}_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q,$$

where

- $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$;
- $\boldsymbol{\gamma}_q = (\gamma_q(1), \ldots, \gamma_q(p))^\top$ with $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t, \Delta_q \varepsilon_{t-\ell})$;

Estimator, first stage

Yule-Walker equations yield

$$\boldsymbol{\Gamma}_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q,$$

where

- $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$;
- $\boldsymbol{\gamma}_q = (\gamma_q(1), \ldots, \gamma_q(p))^\top$ with $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t, \Delta_q \varepsilon_{t-\ell})$;
- and $\boldsymbol{\Gamma}_q$ is the $p \times p$ covariance matrix $\boldsymbol{\Gamma}_q = (\gamma_q(i-j) : 1 \leq i, j \leq p)$.

Estimator, first stage

Yule-Walker equations yield

$$\boldsymbol{\Gamma}_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q,$$

where

- $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$;
- $\boldsymbol{\gamma}_q = (\gamma_q(1), \ldots, \gamma_q(p))^\top$ with $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t, \Delta_q \varepsilon_{t-\ell})$;
- and $\boldsymbol{\Gamma}_q$ is the $p \times p$ covariance matrix $\boldsymbol{\Gamma}_q = (\gamma_q(i-j) : 1 \leq i, j \leq p)$.

Note

$\boldsymbol{\Gamma}_q \boldsymbol{a} \approx \boldsymbol{\gamma}_q$ for large values of q.

Estimator, first stage

Yule-Walker equations yield

$$\boldsymbol{\Gamma}_q \boldsymbol{a} = \boldsymbol{\gamma}_q + \nu^2 \boldsymbol{c}_q,$$

where

- $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ are the coefficients from the MA(∞) expansion of $\{\varepsilon_t\}$;
- $\boldsymbol{\gamma}_q = (\gamma_q(1), \ldots, \gamma_q(p))^\top$ with $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q \varepsilon_t, \Delta_q \varepsilon_{t-\ell})$;
- and $\boldsymbol{\Gamma}_q$ is the $p \times p$ covariance matrix $\boldsymbol{\Gamma}_q = (\gamma_q(i-j) : 1 \leq i, j \leq p)$.

Note

$\boldsymbol{\Gamma}_q \boldsymbol{a} \approx \boldsymbol{\gamma}_q$ for large values of q.

We construct the first-stage estimator by

$$\widetilde{\boldsymbol{a}}_q = \widehat{\boldsymbol{\Gamma}}_q^{-1} \widehat{\boldsymbol{\gamma}}_q,$$

where $\widehat{\boldsymbol{\Gamma}}_q$ and $\widehat{\boldsymbol{\gamma}}_q$ are constructed from the sample autocovariances $\widehat{\gamma}_q(\ell) = (T-q)^{-1} \sum_{t=q+\ell+1}^{T} \Delta_q Y_{t,T} \Delta_q Y_{t-\ell,T}$.

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Solution:

- Compute estimators $\widetilde{c}_k$ of c_k based on $\widetilde{\boldsymbol{a}}_q$.

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Solution:

- Compute estimators $\widetilde{c}_k$ of c_k based on $\widetilde{\boldsymbol{a}}_q$.
- Estimate the innovation variance ν^2 by $\widetilde{\nu}^2 = (2T)^{-1} \sum_{t=p+1}^{T} \widetilde{r}_{t,T}^2$, where $\widetilde{r}_{t,T} = \Delta_1 Y_{t,T} - \sum_{j=1}^{p} \widetilde{a}_j \Delta_1 Y_{t-j,T}$.

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Solution:

- Compute estimators $\widetilde{c}_k$ of c_k based on $\widetilde{\boldsymbol{a}}_q$.
- Estimate the innovation variance ν^2 by $\widetilde{\nu}^2 = (2T)^{-1} \sum_{t=p+1}^{T} \widetilde{r}_{t,T}^2$, where $\widetilde{r}_{t,T} = \Delta_1 Y_{t,T} - \sum_{j=1}^{p} \widetilde{a}_j \Delta_1 Y_{t-j,T}$.
- Estimate $\boldsymbol{a}$ by

$$\widehat{\boldsymbol{a}}_r = \widehat{\boldsymbol{\Gamma}}_r^{-1} (\widehat{\boldsymbol{\gamma}}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r).$$

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Solution:

- Compute estimators $\widetilde{c}_k$ of c_k based on $\widetilde{\boldsymbol{a}}_q$.
- Estimate the innovation variance ν^2 by $\widetilde{\nu}^2 = (2T)^{-1} \sum_{t=p+1}^{T} \widetilde{r}_{t,T}^2$, where $\widetilde{r}_{t,T} = \Delta_1 Y_{t,T} - \sum_{j=1}^{p} \widetilde{a}_j \Delta_1 Y_{t-j,T}$.
- Estimate $\boldsymbol{a}$ by
$$\widehat{\boldsymbol{a}}_r = \widehat{\boldsymbol{\Gamma}}_r^{-1} (\widehat{\gamma}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r).$$
- Average the estimators $\widehat{\boldsymbol{a}}_r$: $\widehat{\boldsymbol{a}} = \frac{1}{\bar{r}} \sum_{r=1}^{\bar{r}} \widehat{\boldsymbol{a}}_r$.

Estimator, second stage

Problem

If the trend m is pronounced, the estimator $\widetilde{\boldsymbol{a}}_q$ will have a strong bias.

Solution:

- Compute estimators $\widetilde{c}_k$ of c_k based on $\widetilde{\boldsymbol{a}}_q$.
- Estimate the innovation variance ν^2 by $\widetilde{\nu}^2 = (2T)^{-1} \sum_{t=p+1}^{T} \widetilde{r}_{t,T}^2$, where $\widetilde{r}_{t,T} = \Delta_1 Y_{t,T} - \sum_{j=1}^{p} \widetilde{a}_j \Delta_1 Y_{t-j,T}$.
- Estimate $\boldsymbol{a}$ by
$$\widehat{\boldsymbol{a}}_r = \widehat{\boldsymbol{\Gamma}}_r^{-1} (\widehat{\boldsymbol{\gamma}}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r).$$
- Average the estimators $\widehat{\boldsymbol{a}}_r$: $\widehat{\boldsymbol{a}} = \frac{1}{\bar{r}} \sum\limits_{r=1}^{\bar{r}} \widehat{\boldsymbol{a}}_r$.
- Estimate the long-run variance σ^2 by
$$\widehat{\sigma}^2 = \frac{\widehat{\nu}^2}{(1 - \sum_{j=1}^{p} \widehat{a}_j)^2}.$$

Motivation for the estimator

If $\{\varepsilon_t\}$ is an AR(p) process, then the time series $\{\Delta_q \varepsilon_t\}$ of the differences $\Delta_q \varepsilon_t = \varepsilon_t - \varepsilon_{t-q}$ is an ARMA(p, q) process of the form

$$\Delta_q \varepsilon_t - \sum_{j=1}^{p} a_j \Delta_q \varepsilon_{t-j} = \eta_t - \eta_{t-q}.$$

Motivation for the estimator

If $\{\varepsilon_t\}$ is an AR(p) process, then the time series $\{\Delta_q \varepsilon_t\}$ of the differences $\Delta_q \varepsilon_t = \varepsilon_t - \varepsilon_{t-q}$ is an ARMA(p, q) process of the form

$$\Delta_q \varepsilon_t - \sum_{j=1}^{p} a_j \Delta_q \varepsilon_{t-j} = \eta_t - \eta_{t-q}.$$

Then $\Delta_q Y_{t,T} = Y_{t,T} - Y_{t-q,T}$ is approximately an ARMA(p, q) process.

Theoretical properties of the estimator

Performance:

- Our estimator $\widehat{\boldsymbol{a}}$ produces accurate estimation results even when the AR polynomial $A(z) = 1 - \sum_{j=1}^{p} a_j z^j$ has a root close to the unit circle.

Theoretical properties of the estimator

Performance:

- Our estimator $\widehat{\boldsymbol{a}}$ produces accurate estimation results even when the AR polynomial $A(z) = 1 - \sum_{j=1}^{p} a_j z^j$ has a root close to the unit circle.
- Our pilot estimator $\widetilde{\boldsymbol{a}}_q$ tends to have a substantial bias when the trend m is pronounced. Our estimator $\widehat{\boldsymbol{a}}$ reduces this bias considerably.

Theoretical properties of the estimator

Performance:

- Our estimator $\widehat{\boldsymbol{a}}$ produces accurate estimation results even when the AR polynomial $A(z) = 1 - \sum_{j=1}^{p} a_j z^j$ has a root close to the unit circle.
- Our pilot estimator $\widetilde{\boldsymbol{a}}_q$ tends to have a substantial bias when the trend m is pronounced. Our estimator $\widehat{\boldsymbol{a}}$ reduces this bias considerably.

Proposition

Our estimators $\widetilde{\boldsymbol{a}}_q$, $\widehat{\boldsymbol{a}}$ and $\widehat{\sigma}^2$ are $\sqrt{T}$-consistent.

Idea behind the additive correction

Consider the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \right|$$

under the null hypothesis $H_0 : m = 0$ and under simplifying assumptions:

- the errors ε_i are i.i.d. normally distributed;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

Idea behind the additive correction

Consider the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \right|$$

under the null hypothesis $H_0 : m = 0$ and under simplifying assumptions:

- the errors ε_i are i.i.d. normally distributed;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1\leq l\leq L} \max_{1\leq k\leq 1/2h_l} \left| \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} \right|$$

Idea behind the additive correction

Consider the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \right|$$

under the null hypothesis $H_0 : m = 0$ and under simplifying assumptions:

- the errors ε_i are i.i.d. normally distributed;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \le k \le 1/2h_l, 1 \le l \le L\}$.

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1\le l\le L} \max_{1\le k\le 1/2h_l} \left| \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} \right|$$

Idea behind the additive correction

Consider the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \left|\frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}}\right|$$

under the null hypothesis $H_0 : m = 0$ and under simplifying assumptions:

- the errors ε_i are i.i.d. normally distributed;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1\leq l\leq L} \max_{1\leq k\leq 1/2h_l} \left|\frac{\widehat{\psi}_T(u_k,h_l)}{\sigma}\right|$$

$\Rightarrow \quad \max_k \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} = \sqrt{2\log(1/2h_l)} + o_P(1) \to \infty$ as $h \to 0$ and the stochastic behavior of $\widehat{\Psi}_{T,\text{uncorrected}}$ is dominated by $\frac{\widehat{\psi}_T(u_k,h_l)}{\sigma}$ for small bandwidths h_l. Go back

Idea behind the additive correction

Consider the uncorrected statistic

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{(u,h)\in\mathcal{G}_T} \left| \frac{\widehat{\psi}_T(u,h)}{\widehat{\sigma}} \right|$$

under the null hypothesis $H_0 : m = 0$ and under simplifying assumptions:

- the errors ε_i are i.i.d. normally distributed;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

$$\widehat{\Psi}_{T,\text{uncorrected}} = \max_{1\leq l\leq L} \max_{1\leq k\leq 1/2h_l} \left| \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} \right|$$

$\Rightarrow$ $\max_k \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} = \sqrt{2\log(1/2h_l)} + o_P(1) \to \infty$ as $h \to 0$ and the stochastic behavior of $\widehat{\Psi}_{T,\text{uncorrected}}$ is dominated by $\frac{\widehat{\psi}_T(u_k,h_l)}{\sigma}$ for small bandwidths h_l. Go back

Idea behind $\hat{\sigma}$

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.

Idea behind $\hat{\sigma}$

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,

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

Idea behind $\hat{\sigma}$

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

Idea behind a_k and b_k

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

- Traditional approach: $c_T(\alpha) = c_{ijk,T}(\alpha)$ for all (i, j, k).

Idea behind a_k and b_k

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

- Traditional approach: $c_T(\alpha) = c_{ijk,T}(\alpha)$ for all (i, j, k).
- A more modern approach: $c_{ijk,T}(\alpha)$ depend on the length h_k of the time interval (Dümbgen and Spokoiny (2001)). In our context:

  $$c_{ijk,T}(\alpha) = c_T(\alpha, h_k) := b_k + q_T(\alpha)/a_k,$$

  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 $q_T(\alpha)$ is the $(1-\alpha)$-quantile of the statistic

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

  in order to ensure control of the FWER at level α.

Go back