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

Marina Khismatullina Michael Vogt

01/10/2020

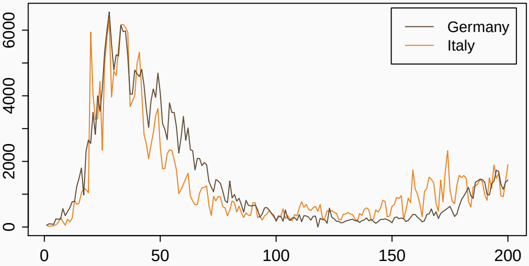
Table of contents

Introduction

Motivation

Research question:

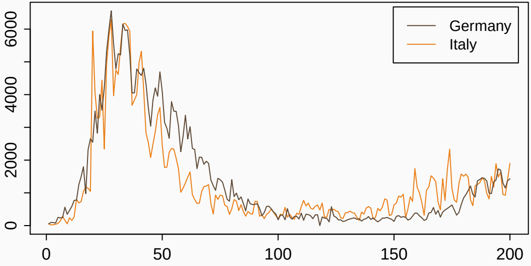
How do outbreak patterns of COVID-19 compare across countries?



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 for $1 \leq i \leq n$

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

Model

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

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

where

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

Literature

Curve comparisons

-

Literature

Curve comparisons

-

Studies of COVID-19

- SEIR models

The multiscale method

Testing

Let $\mathcal{F} = \{\mathcal{I}_k \subset [0,1] : 1 \leq k \leq K\}$ 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 then

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

Testing

Let $\mathcal{F} = \{\mathcal{I}_k \subset [0,1] : 1 \leq k \leq K\}$ 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 then

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

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 an interval $\mathcal{I}_k$ and a pair of time series i and j we construct the kernel averages

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

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

Test statistic

For an interval $\mathcal{I}_k$ and a pair of time series i and j we construct the kernel averages

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

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

Under certain assumptions,

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

Test statistic

For an interval $\mathcal{I}_k$ and a pair of time series i and j we construct the kernel averages

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

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

Under certain assumptions,

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

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

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

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 follows

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

Test procedure

Testing problem:

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

Test procedure

Testing problem:

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

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_t are independent standard normal random variables;

$q_{n,T}(\alpha)$ is $(1-\alpha)$ quantile of $\Phi_{n,T}$.

Test procedure

Testing problem:

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

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_t are independent standard normal random variables;

$q_{n,T}(\alpha)$ is $(1-\alpha)$ quantile of $\Phi_{n,T}$.

Test procedure

For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_{n,T} > q_{n,T}(\alpha)$.

Theoretical properties

Proposition

Supose that $\mathcal{E}_i$ are independent across i and satisfy $\mathcal{C}1 - \mathcal{C}2$ for each i. Under our remaining assumptions and under $H_0 : m_1 = m_2 = \ldots = m_n$ it holds that

$$\mathrm{P}\big(\widehat{\Psi}_{n,T} \le q_{n,T}(\alpha)\big) = (1-\alpha) + o(1).$$

Theoretical properties

Proposition

Supose that $\mathcal{E}_i$ are independent across i and satisfy $\mathcal{C}1 - \mathcal{C}2$ for each i. Under our remaining assumptions and under $H_0 : m_1 = m_2 = \ldots = m_n$ it holds that

$$\mathrm{P}\big(\widehat{\Psi}_{n,T} \le q_{n,T}(\alpha)\big) = (1-\alpha) + o(1).$$

Proposition

Let the conditions of previous proposition be satisfied. Under local alternatives we have

$$\mathrm{P}\big(\widehat{\Psi}_{n,T} \le q_{n,T}(\alpha)\big) = o(1).$$

Test procedure

Gaussian version of the test statistic:

$$\Phi_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi_T(u,h)}{\sigma} \Big| - \lambda(h) \Big\},$$

where

- $\phi_T(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\, \sigma Z_t$;
- Z_t are independent standard normal random variables;
- $q_T(\alpha)$ is $(1-\alpha)$ quantile of Φ_T.

Test procedure

Gaussian version of the test statistic:

$$\Phi_T = \max_{(u,h)\in\mathcal{G}_T} \Big\{ \Big| \frac{\phi_T(u,h)}{\sigma} \Big| - \lambda(h) \Big\},$$

where

- $\phi_T(u,h) = \sum_{t=1}^{T} w_{t,T}(u,h)\, \sigma Z_t$;
- Z_t are independent standard normal random variables;
- $q_T(\alpha)$ is $(1-\alpha)$ quantile of Φ_T.

Test procedure

For a given significance level $\alpha \in (0,1)$, we reject H_0 if $\widehat{\Psi}_T > q_T(\alpha)$.

Theoretical properties

Assumptions

$\mathcal{C}1$ The variables ε_t are weakly dependent.

Assumptions

$\mathcal{C}1$ The variables ε_t are weakly dependent.

$\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.

Assumptions

- $\mathcal{C}1$ The variables ε_t are weakly dependent.
- $\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.
- $\mathcal{C}3$ Standard assumptions on the kernel function K.

Assumptions

- $\mathcal{C}1$ The variables ε_t are weakly dependent.
- $\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.
- $\mathcal{C}3$ Standard assumptions on the kernel function K.
- $\mathcal{C}4$ Assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$.

Assumptions

- $\mathcal{C}1$ The variables ε_t are weakly dependent.
- $\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.
- $\mathcal{C}3$ Standard assumptions on the kernel function K.
- $\mathcal{C}4$ Assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$.
- $\mathcal{C}5$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

Assumptions

$\mathcal{C}1$ The variables ε_t are weakly dependent.

$\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.

$\mathcal{C}3$ Standard assumptions on the kernel function K.

$\mathcal{C}4$ Assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$.

$\mathcal{C}5$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

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

Assumptions

$\mathcal{C}1$ The variables ε_t are weakly dependent.

$\mathcal{C}2$ It holds that $\|\varepsilon_t\|_q < \infty$ for some $q > 4$.

$\mathcal{C}3$ Standard assumptions on the kernel function K.

$\mathcal{C}4$ Assume that $\widehat{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ with $\rho_T = o(1/\log T)$.

$\mathcal{C}5$ $|\mathcal{G}_T| = O(T^\theta)$ for some arbitrarily large but fixed constant $\theta > 0$.

$\mathcal{C}6$ $h_{\min} \gg T^{-(1-\frac{2}{q})} \log T$ and $h_{\max} = o(1)$.

Theoretical properties

Proposition

Under our assumptions and under $H_0 : m' = 0$ it holds that

$$\mathrm{P}\big(\widehat{\Psi}_T \leq q_T(\alpha)\big) = (1-\alpha) + o(1).$$

Theoretical properties

Proposition

Under our assumptions and under $H_0 : m' = 0$ it holds that

$$\mathrm{P}\big(\widehat{\Psi}_T \leq q_T(\alpha)\big) = (1-\alpha) + o(1).$$

Proposition

Under our assumptions and under local alternatives, we have

$$\mathrm{P}\big(\widehat{\Psi}_T \leq q_T(\alpha)\big) = o(1).$$

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.

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

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.

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

- Show that

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

Theoretical properties

Define

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

with

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

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*

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

Theoretical properties

Proposition

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

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

Graphical representation

Minimal intervals

An interval $I_{u,h} \in \Pi_T^+$ is called **minimal** if there is no other interval $I_{u',h'} \in \Pi_T^+$ with $I_{u',h'} \subset I_{u,h}$.

Graphical representation

Minimal intervals

An interval $I_{u,h} \in \Pi_T^+$ is called **minimal** if there is no other interval $I_{u',h'} \in \Pi_T^+$ with $I_{u',h'} \subset I_{u,h}$.

Define

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

Graphical representation

Minimal intervals

An interval $I_{u,h} \in \Pi_T^+$ is called **minimal** if there is no other interval $I_{u',h'} \in \Pi_T^+$ with $I_{u',h'} \subset I_{u,h}$.

Define

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

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

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

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

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

- $\boldsymbol{a} = (a_1, \dots, a_p)$ 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

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

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 $\mathbf{\Gamma}_q$ is the $p \times p$ covariance matrix $\mathbf{\Gamma}_q = (\gamma_q(i-j) : 1 \leq i, j \leq p)$.

Note

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

We construct the first-stage estimator by

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

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

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
$$\widehat{\boldsymbol{a}}_r = \widehat{\boldsymbol{\Gamma}}_r^{-1} (\widehat{\boldsymbol{\gamma}}_r + \widetilde{\nu}^2 \widetilde{\boldsymbol{c}}_r).$$
- Average the estimators $\widehat{\boldsymbol{a}}_r$: $\widehat{\boldsymbol{a}} = \frac{1}{\bar{r}} \sum_{r=1}^{\bar{r}} \widehat{\boldsymbol{a}}_r$.

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 $\widehat{\boldsymbol{a}}_r$: $\widehat{\boldsymbol{a}} = \frac{1}{\bar{r}} \sum_{r=1}^{\bar{r}} \widehat{\boldsymbol{a}}_r$.
- 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.

Clustering, group structure

- The null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is violated.

Clustering, group structure

- The null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is violated.
- There exist sets or groups of time series $G_1, \ldots, G_N$ with $N \leq n$ and $\{1, \ldots, n\} = \dot{\bigcup}_{\ell=1}^{N} G_\ell$ such that for each $1 \leq \ell \leq N$ we have $m_i = g_\ell \quad$ for all $i \in G_\ell$, where g_ℓ are group-specific trend functions.

Clustering, group structure

- The null hypothesis $H_0 : m_1 = m_2 = \ldots = m_n$ is violated.
- There exist sets or groups of time series $G_1, \ldots, G_N$ with $N \leq n$ and $\{1, \ldots, n\} = \dot{\bigcup}_{\ell=1}^{N} G_\ell$ such that for each $1 \leq \ell \leq N$ we have $m_i = g_\ell \quad$ for all $i \in G_\ell$, where g_ℓ are group-specific trend functions.
- For any $\ell \neq \ell'$, the trends $g_{\ell,T}$ and $g_{\ell',T}$ differ in the following sense: There exists $(u, h) \in \mathcal{G}_T$ with $[u - h, u + h] \subseteq [0, 1]$ such that $g_{\ell,T}(w) - g_{\ell',T}(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$ or $g_{\ell',T}(w) - g_{\ell,T}(w) \geq c_T \sqrt{\log T/(Th)}$ for all $w \in [u - h, u + h]$, where $0 < c_T \to \infty$.

Clustering, algorithm

Dissimilarity measure between two sets of time series S and S':

$$\widehat{\Delta}(S, S') = \max_{\substack{i \in S, \\ j \in S'}} \widehat{\Psi}_{ij,T}.$$

Clustering algorithm

Step 0 *(Initialization):* Let $\widehat{G}_i^{[0]} = \{i\}$ denote the i-th singleton cluster for $1 \le i \le n$ and define $\{\widehat{G}_1^{[0]}, \ldots, \widehat{G}_n^{[0]}\}$ to be the initial partition of time series into clusters.

Step r (Iteration): Let $\widehat{G}_1^{[r-1]}, \ldots, \widehat{G}_{n-(r-1)}^{[r-1]}$ be the $n-(r-1)$ clusters from the previous step. Determine the pair of clusters $\widehat{G}_\ell^{[r-1]}$ and $\widehat{G}_{\ell'}^{[r-1]}$ for which

$$\widehat{\Delta}(\widehat{G}_\ell^{[r-1]}, \widehat{G}_{\ell'}^{[r-1]}) = \min_{1 \le k < k' \le n-(r-1)} \widehat{\Delta}(\widehat{G}_k^{[r-1]}, \widehat{G}_{k'}^{[r-1]})$$

and merge them into a new cluster.

Clustering, theoretical properties

The estimator of the number of groups is

$$\widehat{N} = \min \Big\{ r = 1, 2, \ldots \,\Big|\, \max_{1 \le \ell \le r} \widehat{\Delta}\big(\widehat{G}_{\ell}^{[n-r]}\big) \le q_{n,T}(\alpha) \Big\}.$$

Clustering, theoretical properties

The estimator of the number of groups is

$$\widehat{N} = \min \Big\{ r = 1, 2, \dots \Big| \max_{1 \le \ell \le r} \widehat{\Delta}\big(\widehat{G}_\ell^{[n-r]}\big) \le q_{n,T}(\alpha) \Big\}.$$

Proposition

Let the conditions of previous propositions be satisfied. Then

$$\mathrm{P}\Big(\big\{\widehat{G}_1, \dots, \widehat{G}_{\widehat{N}}\big\} = \big\{G_1, \dots, G_N\big\}\Big) \ge (1-\alpha) + o(1)$$

and

$$\mathrm{P}\big(\widehat{N} = N\big) \ge (1-\alpha) + o(1).$$

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;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

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

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;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

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

$\Rightarrow \quad \max_k \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} = \sqrt{2\log(1/2h_l)} + o_P(1) \to \infty$ 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;
- $\widehat{\sigma} = \sigma$;
- $\mathcal{G}_T = \{(u_k, h_l) | u_k = (2k-1)h_l \text{ for } 1 \leq k \leq 1/2h_l, 1 \leq l \leq L\}$.

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

$\Rightarrow \quad \max_k \frac{\widehat{\psi}_T(u_k,h_l)}{\sigma} = \sqrt{2\log(1/2h_l)} + o_P(1) \to \infty$ 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 $\hat{\sigma}$

Go back