Clustering of the epidemic time trends: the case of COVID-19

We consider the following nonparametric regression equation:

$$X_{it} = c_i \lambda_i\Big(\frac{t}{T}\Big) + \varepsilon_{it} \quad \text{with} \quad \varepsilon_{it} = \sigma \sqrt{\lambda_i\Big(\frac{t}{T}\Big)} \eta_{it},$$

where c_i is the country-specific scaling parameter that accounts for the size of the country or population density. We introduce this additional parameter in order to be able to compare countries that differ substantially in terms of the population, i.e. Luxembourg and Russia. In what follows, we present a method that allows researchers to test the hypothesis that the time trends of new COVID-19 cases in different countries are the same up to some scaling parameter and to cluster the countries based on the differences.

For the identification purposes, we need to assume that for each $i \in \mathcal{C}$ we have $\int_0^1 \lambda_i(u) du = 1$. Only then we are able to estimate the scaling parameter c_i. Thus, the testing procedure is as follows.

Step 1

First, we estimate the scaling parameter:

$$\begin{aligned}
\widehat{c_i} &= \frac{1}{T} \sum_{t=1}^{T} X_{it} \\
&= c_i \frac{1}{T} \sum_{t=1}^{T} \lambda_i\Big(\frac{t}{T}\Big) + \sigma \frac{1}{T} \sum_{t=1}^{T} \sqrt{\lambda_i\Big(\frac{t}{T}\Big)} \eta_{it} \\
&= c_i \frac{1}{T} \sum_{t=1}^{T} \lambda_i\Big(\frac{t}{T}\Big) + o_P(1) \\
&= c_i + o_P(1),
\end{aligned}$$

where in the last inequality we used the normalization $\int_0^1 \lambda_i(u) du = 1$. Hence, for any fixed $i \in \mathcal{C}$, $\widehat{c_i}$ is a consistent estimator of c_i.

Step 2

Instead of working with X_{it}, we consider the following variables:

$$\begin{aligned}
X_{it}^* &= \frac{X_{it}}{\frac{1}{T} \sum_{t=1}^{T} X_{it}} \\
&= \frac{c_i}{\widehat{c_i}} \lambda_i\Big(\frac{t}{T}\Big) + \frac{\sigma}{\widehat{c_i}} \sqrt{\lambda_i\Big(\frac{t}{T}\Big)} \eta_{it}.
\end{aligned}$$

A statistic to test the hypothesis $H_0^{(ijk)}$ for a given triple (i,j,k) is then constructed as follows. We work with the following quantity

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

Then

$$\begin{aligned}
\frac{\hat{s}_{ijk,T}}{\sqrt{Th_k}} =& \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(X_{it}^* - X_{jt}^*) \\
=& \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big(\lambda_i\Big(\frac{t}{T}\Big) - \lambda_j\Big(\frac{t}{T}\Big)\Big) + R_1 + R_2,
\end{aligned}$$

where

$$\begin{aligned}
R_1 &= \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big(\Big(\frac{c_i}{\widehat{c}_i} - 1\Big)\lambda_i\Big(\frac{t}{T}\Big) - \Big(\frac{c_j}{\widehat{c}_j} - 1\Big)\lambda_j\Big(\frac{t}{T}\Big)\Big), \\
R_2 &= \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big(\frac{\sigma}{\widehat{c}_i}\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it} - \frac{\sigma}{\widehat{c}_j}\sqrt{\lambda_j\Big(\frac{t}{T}\Big)}\eta_{jt}\Big).
\end{aligned}$$

Since $\widehat{c}_i = c_i + o_P(1)$ and $0 \le \sum_{t=1}^{T} \mathbf{1}\big(\frac{t}{T} \in \mathcal{I}_k\big)\lambda_i\big(\frac{t}{T}\big) \le h_k \lambda_{max}$, we have

$$\begin{aligned}
|R_1| \le& \Big|\frac{c_i}{\widehat{c}_i} - 1\Big| \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\lambda_i\Big(\frac{t}{T}\Big) + \Big|\frac{c_j}{\widehat{c}_j} - 1\Big| \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\lambda_j\Big(\frac{t}{T}\Big), \\
&\le o_P(1) \cdot \frac{\lambda_{max}}{T} + o_P(1) \cdot \frac{\lambda_{max}}{T} = o_P\Big(\frac{1}{T}\Big). \qquad (0.1)
\end{aligned}$$

Furthermore, applying the law of large numbers, we get:

$$\frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it} = o_P(1).$$

Hence, if we uniformly bound the scaling parameters away from 0, i.e. $\exists\, c_{min}$ such that for all $i \in \mathcal{C}$ we have $0 < c_{min} \le c_i$, we can use the fact that $\frac{\sigma}{\widehat{c}_i} = O_P(1)$ to get that

$$\begin{aligned}
R_2 &= \frac{\sigma}{\widehat{c}_i} \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\sqrt{\lambda_i\Big(\frac{t}{T}\Big)}\eta_{it} - \frac{\sigma}{\widehat{c}_j} \frac{1}{Th_k} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\sqrt{\lambda_j\Big(\frac{t}{T}\Big)}\eta_{jt} \\
&= o_P(1). \qquad (0.2)
\end{aligned}$$

Combining (0.1) and (0.2) together, we get $\hat{s}_{ijk,T}/\sqrt{Th_k} = (Th_k)^{-1} \sum_{t=1}^{T} \mathbf{1}(t/T \in \mathcal{I}_k)\{\lambda_i(t/T) -$

$\lambda_j(t/T)\} + o_p(1)$ for any fixed pair of countries (i,j). Hence, the statistic $\hat{s}_{ijk,T}/\sqrt{Th_k}$ estimates the average distance between the functions λ_i and λ_j on the interval $\mathcal{I}_k$. The variance of $\hat{s}_{ijk,T}$ can not be easily calculated:

$$\begin{aligned}
\mathrm{Var}(\hat{s}_{ijk,T}) =& \frac{1}{Th_k}\mathrm{Var}\Big(\sum_{t=1}^{T}\mathbf{1}\Big(\frac{t}{T}\in\mathcal{I}_k\Big)(X_{it}^*-X_{jt}^*)\Big)\\
=& \frac{1}{Th_k}\mathrm{Var}\Big(\sum_{t=1}^{T}\mathbf{1}\Big(\frac{t}{T}\in\mathcal{I}_k\Big)X_{it}^*\Big) + \frac{1}{Th_k}\mathrm{Var}\Big(\sum_{t=1}^{T}\mathbf{1}\Big(\frac{t}{T}\in\mathcal{I}_k\Big)X_{jt}^*\Big)\\
=& \frac{1}{Th_k}\mathrm{Var}\bigg(\frac{\sum_{t=1}^{T}\mathbf{1}\big(\frac{t}{T}\in\mathcal{I}_k\big)X_{it}}{\frac{1}{T}\sum_{t=1}^{T}X_{it}}\bigg) + \frac{1}{Th_k}\mathrm{Var}\bigg(\frac{\sum_{t=1}^{T}\mathbf{1}\big(\frac{t}{T}\in\mathcal{I}_k\big)X_{jt}}{\frac{1}{T}\sum_{t=1}^{T}X_{jt}}\bigg),
\end{aligned}$$

hence, we 'normalize' $\hat{s}_{ijk,T}$ intuitively by dividing it by the following value:

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

Instead look at the bootstrap!

Normalizing the statistic $\hat{s}_{ijk,T}$ by the estimator $\hat{\nu}_{ijk,T}$ yields the expression

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

which serves as our test statistic of the hypothesis $H_0^{(ijk)}$. For later reference, we additionally introduce the statistic

$$\hat{\psi}_{ijk,T}^0 = \frac{\sum_{t=1}^{T}\mathbf{1}(\frac{t}{T}\in\mathcal{I}_k)\Big((\frac{c_i}{\hat{c}_i}-\frac{c_j}{\hat{c}_j})\overline{\lambda}_{ij}+\overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\frac{\sigma}{\hat{c}_i}\eta_{it}-\frac{\sigma}{\hat{c}_j}\eta_{jt})\Big)}{\hat{\sigma}\{\sum_{t=1}^{T}\mathbf{1}(\frac{t}{T}\in\mathcal{I}_k)(X_{it}^*+X_{jt}^*)\}^{1/2}} \tag{0.3}$$

with $\overline{\lambda}_{ij}(u) = \{\lambda_i(u)+\lambda_j(u)\}/2$, which is identical to $\hat{\psi}_{ijk,T}$ under $H_0^{(ijk)}$.

0.1 Construction of the test

Our multiscale test is carried out as follows: For a given significance level $\alpha\in(0,1)$ and each $(i,j,k)\in\mathcal{M}$, we reject $H_0^{(ijk)}$ if

$$|\hat{\psi}_{ijk,T}| > c_{ijk,T}(\alpha),$$

where $c_{ijk,T}(\alpha)$ is the critical value for the (i,j,k)-th test problem. The critical values $c_{ijk,T}(\alpha)$ are chosen such that the familywise error rate (FWER) is controlled at level α, which is defined as the probability of wrongly rejecting $H_0^{(ijk)}$ for at least one (i,j,k). More formally speaking, for a given significance level $\alpha \in (0,1)$, the FWER is

$$\begin{aligned}
\text{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{ijk,T}(\alpha)\Big) \\
&= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \le c_{ijk,T}(\alpha)\Big) \\
&= 1 - \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} |\hat{\psi}_{ijk,T}| \le c_{ijk,T}(\alpha)\Big),
\end{aligned}$$

where $\mathcal{M}_0 \subseteq \mathcal{M}$ is the set of triples (i,j,k) for which $H_0^{(ijk)}$ holds true. As before, the critical values are chosen as

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

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 the quantity $q_T(\alpha)$ is determined by the following consideration: Since

$$\begin{aligned}
\text{FWER}(\alpha) &= \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_T(\alpha, h_k)\Big) \\
&= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| \le c_T(\alpha, h_k)\Big) \\
&= 1 - \mathbb{P}\Big(\forall (i,j,k) \in \mathcal{M}_0 : a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) \le q_T(\alpha)\Big) \\
&= 1 - \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} a_k\big(|\hat{\psi}_{ijk,T}| - b_k\big) \le q_T(\alpha)\Big),
\end{aligned}$$

we need to choose the quantity $q_T(\alpha)$ as the $(1-\alpha)$-quantile of the statistic

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

in order to ensure control of the FWER at level α. As the quantiles $q_T(\alpha)$ are not known in practice, we cannot compute the critical values $c_T(\alpha, h_k)$ exactly in practice but need to approximate them. This can be achieved as follows: Under appropriate regularity conditions, it can be shown that ???

$$\hat{\psi}^0_{ijk,T} = \frac{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)\Big(\big(\frac{c_i}{\hat{c}_i} - \frac{c_j}{\hat{c}_j}\big)\overline{\lambda}_{ij} + \overline{\lambda}_{ij}^{1/2}(\frac{t}{T})(\frac{\sigma}{\hat{c}_i}\eta_{it} - \frac{\sigma}{\hat{c}_j}\eta_{jt})\Big)}{\hat{\sigma}\{\sum_{t=1}^T \mathbf{1}(\frac{t}{T} \in \mathcal{I}_k)(X^*_{it} + X^*_{jt})\}^{1/2}}$$

$$\approx \frac{1}{\sqrt{2Th_k}} \sum_{t=1}^{T} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big) \Big\{ \frac{\eta_{it}}{\hat{c}_i} - \frac{\eta_{jt}}{\hat{c}_j} \Big\}.$$

In what follows, we will be using the Gaussian version $\phi_{ijk,T}$ of the statistic displayed above:

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

where Z_{it} are independent standard normal random variables for $1 \leq t \leq T$ and $1 \leq i \leq n$. However, since the variance of $\hat{\psi}^0_{ijk,T}$ is not easy to calculate and we need the variances of $\hat{\psi}^0_{ijk,T}$ and $\phi_{ijk,T}$ to be equal, we can not provide the exact formula for $d_{i,T}$. We regard $\phi_{ijk,T}$ as an auxiliary test statistic with unknown distribution, which will be then approximated by bootstrap. For now, the statistic

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

can be regarded as a Gaussian version of the statistic $\hat{\Psi}_T$. Further in this section, we will show how that the critical values of Φ_T and the unknown quantiles $q_T(\alpha)$ can be approximated using a multiplier bootstrap by the $\Psi_{\text{bootstrap},T}$ and its respective $(1-\alpha)$-quantile $\widehat{q}_{T,\text{bootstrap}}(\alpha)$.

To summarize, we propose the following procedure to simultaneously test the hypothesis $H_0^{(ijk)}$ for all $(i,j,k) \in \mathcal{M}$ at the significance level $\alpha \in (0,1)$:

$$\text{For each } (i,j,k) \in \mathcal{M}, \text{ reject } H_0^{(ijk)} \text{ if } |\hat{\psi}_{ijk,T}| > c_{T,\text{bootstrap}}(\alpha, h_k), \tag{0.4}$$

where $c_{T,\text{bootstrap}}(\alpha, h_k) = b_k + \widehat{q}_{T,\text{bootstrap}}(\alpha)/a_k$ with $a_k = \{\log(e/h_k)\}^{1/2} / \log\log(e^e/h_k)$ and $b_k = \sqrt{2\log(1/h_k)}$.

0.2 Proof strategy

Here we outline the proof strategy, which can be divided into several major steps.

Step 1. Let $\hat{\Psi}_T = \max_{(i,j,k) \in \mathcal{M}} a_k (|\hat{\psi}^0_{ijk,T}| - b_k)$ with $\hat{\psi}^0_{ijk,T}$ introduced in (0.3) and define $\Psi_T = \max_{(i,j,k) \in \mathcal{M}} a_k (|\psi^0_{ijk,T}| - b_k)$ with

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

To start with, we prove that

$$\left|\hat{\Psi}_T - \Psi_T\right| = o_p(r_T), \tag{0.5}$$

where $\{r_T\}$ is any null sequence that converges more slowly to zero than ???

Need to prove that. It is possible that the exact formula of Ψ_T will change.

Step 2. We next prove that

$$\sup_{q\in\mathbb{R}} \Big|\mathbb{P}\big(\Psi_T \le q\big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| = o(1). \tag{0.6}$$

To do so, we rewrite the statistics Ψ_T and Φ_T as follows: Define

$$V_t^{(ijk)} = V_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}}\mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big(\frac{\eta_{it}}{\hat{c}_i} - \frac{\eta_{jt}}{\hat{c}_j}\Big)$$

for $(i,j,k) \in \mathcal{M}$ and let $\boldsymbol{V}_t = (V_t^{(ijk)} : (i,j,k) \in \mathcal{M})$ be the p-dimensional random vector with the entries $V_t^{(ijk)}$. With this notation, we get that $\psi^0_{ijk,T} = T^{-1/2}\sum_{t=1}^T V_t^{(ijk)}$ and thus

$$\begin{aligned}
\Psi_T &= \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\psi^0_{ijk,T}| - b_k\big) \\
&= \max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T V_t^{(ijk)}\Big| - b_k\Big\}.
\end{aligned}$$

Analogously, we define

$$W_t^{(ijk)} = W_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}}\mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)(d_{i,T}Z_{it} - d_{j,T}Z_{jt})$$

with Z_{it} i.i.d. standard normal and let $\boldsymbol{W}_t = (W_t^{(ijk)} : (i,j,k) \in \mathcal{M})$. The vector $\boldsymbol{W}_t$ is a Gaussian version of $\boldsymbol{V}_t$ with the same mean and variance. In particular, $\mathbb{E}[\boldsymbol{W}_t] = \mathbb{E}[\boldsymbol{V}_t] = 0$ and $\mathbb{E}[\boldsymbol{W}_t\boldsymbol{W}_t^\top] = \mathbb{E}[\boldsymbol{V}_t\boldsymbol{V}_t^\top]$ (this we achieve by introducing $d_{i,T}$). Similarly as before, we can write $\phi_{ijk,T} = T^{-1/2}\sum_{t=1}^T W_t^{(ijk)}$ and

$$\begin{aligned}
\Phi_T &= \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\phi_{ijk,T}| - b_k\big) \\
&= \max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T W_t^{(ijk)}\Big| - b_k\Big\}.
\end{aligned}$$

For any $q \in \mathbb{R}$, it holds that

$$\begin{aligned}
\mathbb{P}\big(\Psi_T \le q\big) &= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k \Big\{ \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \Big| - b_k \Big\} \le q \Big) \\
&= \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T V_t^{(ijk)} \Big| \le c_{ijk}(q) \text{ for all } (i,j,k) \in \mathcal{M} \Big) \\
&= \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{V}_t \Big| \le \boldsymbol{c}(q) \Big),
\end{aligned}$$

where $\boldsymbol{c}(q) = (c_{ijk}(q) : (i,j,k) \in \mathcal{M})$ is the $\mathbb{R}^p$-vector with the entries $c_{ijk}(q) = q/a_k + b_k$, we use the notation $|v| = (|v_1|, \ldots, |v_p|)^\top$ for vectors $v \in \mathbb{R}^p$ and the inequality $v \le w$ is to be understood componentwise for $v, w \in \mathbb{R}^p$. Analogously, we have

$$\mathbb{P}\big(\Phi_T \le q\big) = \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{W}_t \Big| \le \boldsymbol{c}(q) \Big).$$

With this notation at hand, we can make use of Proposition 2.1 from Chernozhukov et al. (2017). In our context, this proposition can be stated as follows:

Proposition A.1. *Assume that*

(a) $T^{-1} \sum_{t=1}^T \mathbb{E}(V_t^{(ijk)})^2 \ge \delta > 0$ *for all* $(i,j,k) \in \mathcal{M}$.

(b) $T^{-1} \sum_{t=1}^T \mathbb{E}[|V_t^{(ijk)}|^{2+r}] \le B_T^r$ *for all* $(i,j,k) \in \mathcal{M}$ *and* $r = 1,2$, *where* $B_T \ge 1$ *are constants that may tend to infinity as* $T \to \infty$.

(c) $\mathbb{E}[\{\max_{(i,j,k)\in\mathcal{M}} |V_t^{(ijk)}|/B_T\}^\theta] \le 2$ *for all* t *and some* $\theta > 4$.

Then

$$\begin{aligned}
\sup_{\boldsymbol{c}\in\mathbb{R}^p} \Big| \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{V}_t \Big| \le \boldsymbol{c} \Big) - \mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{W}_t \Big| \le \boldsymbol{c} \Big) \Big| & \\
\le C \Big\{ \Big(\frac{B_T^2 \log^7(pT)}{T} \Big)^{1/6} + \Big(\frac{B_T^2 \log^3(pT)}{T^{1-2/\theta}} \Big)^{1/3} \Big\}, & \qquad (0.7)
\end{aligned}$$

where C *depends only on* δ *and* θ.

We need to check that with our choice of the test statstics, the assumptions (a)–(c) are satisfied. For which B_T?

Hence, Proposition A.2 yields that

$$\sup_{\boldsymbol{c}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{V}_t\Big| \le \boldsymbol{c}\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{W}_t\Big| \le \boldsymbol{c}\Big)\Big| = o(1),$$

which in turn implies (0.6).

Step 3. Now we have the problem that $d_{i,T}$ are unknown. This means that the quantile $q_T(\alpha)$ of Φ_T are not known and can not be approximated by usual Monte Carlo simulations. We need to find another way of approximating them, for example, multiplier bootstrap from Chernozhukov et al. (2017).

We next prove that for $\alpha \in (0, e^{-1})$ we can claim with probability $1-\alpha$ that

$$\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\tilde{\Psi}_T \le q \big| X_{it}) - \mathbb{P}\big(\Phi_T \le q\big)\Big| = o(1), \tag{0.8}$$

where $\tilde{\Psi}_T$ will be defined further.

We have already rewritten the statistics Φ_T as follows:

$$\begin{aligned}\Phi_T &= \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\phi_{ijk,T}| - b_k\big)\\ &= \max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T W_t^{(ijk)}\Big| - b_k\Big\}\end{aligned}$$

and for any $q \in \mathbb{R}$, it holds that

$$\mathbb{P}\big(\Phi_T \le q\big) = \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{W}_t\Big| \le \boldsymbol{c}(q)\Big),$$

where $\boldsymbol{c}(q) = (c_{ijk}(q) : (i,j,k) \in \mathcal{M})$ is the $\mathbb{R}^p$-vector with the entries $c_{ijk}(q) = q/a_k + b_k$.

In order to approximate the unknown distribution of Φ_T by the multiplier bootstrap, we introduce auxiliary test statistics

$$\tilde{\Psi}_T = \max_{(i,j,k)\in\mathcal{M}} a_k\big(|\tilde{\psi}_{ijk,T}| - b_k\big)$$

and

$$\tilde{\psi}_{ijk,T} = \frac{1}{\sqrt{2Th_k}}\sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T}\in\mathcal{I}_k\Big)e_t\Bigg(\Big(\frac{\eta_{it}}{\hat{c}_i} - \frac{\eta_{jt}}{\hat{c}_j}\Big) - \frac{1}{T}\sum_{t=1}^T \mathbf{1}\Big(\frac{t}{T}\in\mathcal{I}_k\Big)\Big(\frac{\eta_{it}}{\hat{c}_i} - \frac{\eta_{jt}}{\hat{c}_j}\Big)\Bigg).$$

with $e_1, \ldots, e_T$ being i.i.d. standard normal random variables independent of $\{Z_{it}|i,t\}$ and $\{X_{it}|i,t\}$. We will use the conditional distribution of $\tilde{\Psi}_T$ given the data as approximation of the distribution of Φ_T.

We proceed as follows. As before,

$$V_t^{(ijk)} = V_{t,T}^{(ijk)} := \sqrt{\frac{T}{2Th_k}} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big\{\frac{\eta_{it}}{\hat{c}_i} - \frac{\eta_{jt}}{\hat{c}_j}\Big\}$$

for $(i,j,k) \in \mathcal{M}$ and let $\boldsymbol{V}_t = (V_t^{(ijk)} : (i,j,k) \in \mathcal{M})$ be the p-dimensional random vector with the entries $V_t^{(ijk)}$. Additionally, define

$$\bar{V}^{(ijk)} = \bar{V}_T^{(ijk)} := \frac{1}{T}\sum_{t=1}^T \sqrt{\frac{T}{2Th_k}} \mathbf{1}\Big(\frac{t}{T} \in \mathcal{I}_k\Big)\Big\{\frac{\eta_{it}}{\hat{c}_i} - \frac{\eta_{jt}}{\hat{c}_j}\Big\}$$

and let $\bar{\boldsymbol{V}} = (\bar{V}^{(ijk)} : (i,j,k) \in \mathcal{M})$ be the p-dimensional random vector with the entries $\bar{V}^{(ijk)}$. We consider the following conditional probability:

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T e_t(\boldsymbol{V}_t - \bar{\boldsymbol{V}})\Big| \le \boldsymbol{c}(q)\Big|\{V_t^{(ijk)}\}\Big) = \\
&= \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T e_t(V_t^{(ijk)} - \bar{V}^{(ijk)})\Big| \le c_{ijk}(q) \text{ for all } (i,j,k) \in \mathcal{M}\Big|\{V_t^{(ijk)}\}\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k\Big\{\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^T e_t(V_t^{(ijk)} - \bar{V}^{(ijk)})\Big| - b_k\Big\} \le q\Big|\{V_t^{(ijk)}\}\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k\Big(\big|\tilde{\psi}_{ijk,T}\big| - b_k\Big) \le q\Big|\{X_{it}\}\Big) \\
&= \mathbb{P}\Big(\tilde{\Psi}_T \le q\Big|\{X_{it}\}\Big)
\end{aligned}$$

We have already checked in the previous step that $\mathbb{E}[\boldsymbol{W}_t] = \mathbb{E}[\boldsymbol{V}_t] = 0$ and $\mathbb{E}[\boldsymbol{W}_t\boldsymbol{W}_t^\top] = \mathbb{E}[\boldsymbol{V}_t\boldsymbol{V}_t^\top]$.

With this notation at hand, we can make use of Corollary 4.2 from Chernozhukov et al. (2017). In our context, this proposition can be stated as follows:

Proposition A.2. *Let $\alpha \in (0, e^{-1})$ be a constant and assume that*

(a) $T^{-1}\sum_{t=1}^T \mathbb{E}(V_t^{(ijk)})^2 \ge \delta > 0$ *for all* $(i,j,k) \in \mathcal{M}$.

(b) $T^{-1}\sum_{t=1}^T \mathbb{E}[|V_t^{(ijk)}|^{2+r}] \le B_T^r$ *for all* $(i,j,k) \in \mathcal{M}$ *and* $r = 1, 2$, *where* $B_T \ge 1$ *are constants that may tend to infinity as* $T \to \infty$.

(c) $\mathbb{E}[\{\max_{(i,j,k)\in\mathcal{M}} |V_t^{(ijk)}|/B_T\}^{\theta}] \le 2$ *for all* t *and some* $\theta > 4$.

Then we have with probability at least $1-\alpha$*,*

$$\begin{aligned}\sup_{\boldsymbol{c}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t(\boldsymbol{V}_t-\bar{\boldsymbol{V}}_t)\Big| \le \boldsymbol{c}(q)\Big|\{V_t^{(ijk)}\}\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\Big| \le \boldsymbol{c}\Big)\Big| \\ \le C\Big\{\Big(\frac{B_T^2\log^5(pT)\log^2(1/\alpha)}{T}\Big)^{1/6} + \Big(\frac{B_T^2\log^3(pT)}{\alpha^{2/\theta}T^{1-2/\theta}}\Big)^{1/3}\Big\},\end{aligned} \tag{0.9}$$

where C *depends only on* δ *and* θ*.*

Again, assumptions (a)–(c) are satisfied if we checked that in the previous step. We just need to determine the rate of B_T such that the RHS is $o(1)$. Hence, Proposition A.2 yields that with probability at least $1-\alpha$

$$\sup_{\boldsymbol{c}\in\mathbb{R}^p}\Big|\mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} e_t(\boldsymbol{V}_t-\bar{\boldsymbol{V}}_t)\Big| \le \boldsymbol{c}(q)\Big|\{V_t^{(ijk)}\}\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{W}_t\Big| \le \boldsymbol{c}\Big)\Big| = o(1),$$

which in turn implies (0.8).

Step 4. Need to rewrite that.

With the help of (0.5) and (0.6), we now show that

$$\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\hat{\Psi}_T \le q\big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| = o(1). \tag{0.10}$$

To start with, the above supremum can be bounded by

$$\begin{aligned}&\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\big(\hat{\Psi}_T \le q\big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| \\ &= \sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \le q + \big\{\Psi_T - \hat{\Psi}_T\big\}\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| \\ &\le \sup_{q\in\mathbb{R}}\max\Big\{\Big|\mathbb{P}\Big(\Psi_T \le q + \big|\Psi_T - \hat{\Psi}_T\big|\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big|, \\ &\qquad\qquad \Big|\mathbb{P}\Big(\Psi_T \le q - \big|\Psi_T - \hat{\Psi}_T\big|\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big|\Big\} \\ &\le \sup_{q\in\mathbb{R}}\max\Big\{\Big|\mathbb{P}\Big(\Psi_T \le q + r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big), \\ &\qquad\qquad \Big|\mathbb{P}\Big(\Psi_T \le q - r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)\Big\} \\ &\le \max_{\ell=0,1}\sup_{q\in\mathbb{R}}\Big|\mathbb{P}\Big(\Psi_T \le q + (-1)^{\ell} r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big)\Big| + \mathbb{P}\Big(\big|\Psi_T - \hat{\Psi}_T\big| > r_T\Big)\end{aligned}$$

$$= \max_{\ell=0,1} \sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Psi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big) \Big| + o(1), \tag{0.11}$$

where the last line is by (0.5). Moreover, for $\ell = 0, 1$,

$$\begin{aligned}
&\sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Psi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big) \Big| \\
&\le \sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Psi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) \Big| \\
&\quad + \sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big) \Big| \\
&= \sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big) \Big| + o(1), \qquad (0.12)
\end{aligned}$$

the last line following from (0.6). Finally, by Nazarov's inequality (**?**, **?** and Lemma A.1 in Chernozhukov et al., 2017), we have that for $\ell = 0, 1$,

$$\begin{aligned}
&\sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Phi_T \le q + (-1)^\ell r_T\Big) - \mathbb{P}\big(\Phi_T \le q\big) \Big| \\
&= \sup_{q\in\mathbb{R}} \Big| \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{W}_t\Big| \le \boldsymbol{c}(q + (-1)^\ell r_T)\Big) - \mathbb{P}\Big(\Big|\frac{1}{\sqrt{T}} \sum_{t=1}^T \boldsymbol{W}_t\Big| \le \boldsymbol{c}(q)\Big) \Big| \\
&\le C \frac{r_T \sqrt{\log(2p)}}{\min_{1\le k\le K} a_k} \le C r_T \sqrt{\log\log T} \sqrt{\log(2p)}, \qquad (0.13)
\end{aligned}$$

where C is a constant that depends only on the parameter δ defined in condition (a) of Proposition A.2 and we have used the fact that $\min_k a_k \ge c/\sqrt{\log\log T}$ for some $c > 0$. Inserting (0.12) and (0.13) into equation (0.11) completes the proof of (0.10).

Step 5. By definition of the quantile $q_{T,\text{Gauss}}(\alpha)$, it holds that $\mathbb{P}(\Phi_T \le q_{T,\text{Gauss}}(\alpha)) \ge 1-\alpha$. As shown in the Supplementary Material, we even have that

$$\mathbb{P}(\Phi_T \le q_{T,\text{Gauss}}(\alpha)) = 1 - \alpha \tag{0.14}$$

for any $\alpha \in (0, 1)$. From this and (0.10), it immediately follows that

$$\mathbb{P}\big(\hat{\Psi}_T \le q_{T,\text{Gauss}}(\alpha)\big) = 1 - \alpha + o(1), \tag{0.15}$$

which in turn implies that

$$\text{FWER}(\alpha) = \mathbb{P}\Big(\exists (i,j,k) \in \mathcal{M}_0 : |\hat{\psi}_{ijk,T}| > c_{T,\text{Gauss}}(\alpha, h_k)\Big)$$

$$
\begin{aligned}
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} a_k \big(|\hat{\psi}_{ijk,T}| - b_k\big) > q_{T,\text{Gauss}}(\alpha)\Big) \\
&= \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}_0} a_k \big(|\hat{\psi}^0_{ijk,T}| - b_k\big) > q_{T,\text{Gauss}}(\alpha)\Big) \\
&\leq \mathbb{P}\Big(\max_{(i,j,k)\in\mathcal{M}} a_k \big(|\hat{\psi}^0_{ijk,T}| - b_k\big) > q_{T,\text{Gauss}}(\alpha)\Big) \\
&= \mathbb{P}\big(\hat{\Psi}_T > q_{T,\text{Gauss}}(\alpha)\big) = \alpha + o(1).
\end{aligned}
$$

This completes the proof of Theorem **??**.

References

Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, **45** 2309–2352.