# On Asymptotic Efficiency of Two-Sample Permutation Tests

Sergey Tarima[a], Viatcheslav Melas[b], Dmitri Salnikov [b]

[a]Medical College of Wisconsin, 8701 Watertown plank Rd., Wauwatosa, WI, USA;
[b]Saint Peterburg State University, Saint Peterburg, Russia

**ARTICLE HISTORY**
Compiled November 3, 2020

**ABSTRACT**
When auto-regression is used to model explosive phase of COVID-19 epidemic, distribution of regression coefficients is heavy tailed and converges to a Cauchy distribution. Permutation tests based on averaged (non-Euclidean) distances show good power properties against many heavy tailed distributions. We found that Pitman's asymptotic efficiency of test statistics based on averaged distances between elements of two samples is equal zero. This finding confusing due excellent performance of such permutation tests in various simulation studies: power was similar to the power of Wilcoxon-Mann-Whitney test for location shift alternatives and was substantially higher for detecting changes in scale. We show that large sample normality of a test statistic is not sufficient to correctly describe its efficiency. There are many test statistics leading to the same permutation test and each of these test statistics has has it own large sample distributions. We showed what test statistic should be used for determining asymptotic efficiency. Overall, we caution researchers to be careful when exploring asymptotic properties of permutation tests. Asymptotic distributions of test statistics maybe insufficient for quantifying asymptotic characteristics of permutation tests.

**KEYWORDS**
U-statistics; heavy tailed distributions; local asympototic properties; contiguous alternatives; energy test

## 1. Introduction

This manuscript starts with a motivating example in Section 1.1, which justifies importance of the problem and the need for comparing power properties of two-samples tests for possibly heavy tailed distributions. The manuscript returns to this example in Section 6. The main focus, however, is not on practical applications but on finite and large sample power properties of two-sample permutation tests. Section 1.2 introduces a few two-sample tests applicable for heavy tailed data.

### 1.1. Motivating example

The COVID-19 pandemic struck every country and no country was fully prepared to deal with it. In every country first weeks were characterized by explosive uncontrolled growth in the number of COVID-19 infections and, consequently, COVID-19 related deaths. The interpretational value of the reported numbers of infections is limited, as it heavily depends on the numbers of performed COVID-19 tests, sensitivity and specificity of tests, and country

CONTACT Email: starima@mcw.edu

specific rules, regulations, and/or guidelines on test administration. One possibility to model the initial explosive phase of the pandemic in $i^{th}$ country is to apply the first order auto-regression for modelling the total number of COVID-19 related deaths on day $t$,

$$Y_{i,t} = \delta_i Y_{i,t-1} + \epsilon_{i,t} \tag{1}$$

where $\epsilon_{i,t}$ are independent and identically distributed errors, $E\epsilon_{i,t} = 0$ and $E\epsilon_{i,t}^2 < \infty$. The partial auto-correlation coefficient $\delta_i$ is a daily multiplicative growth in the number of deaths. A visual inspection of the total numbers of COVID-19 related deaths, downloaded on May 5 2020 from

`https://ourworldindata.org/coronavirus-source-data`

showed that the initial phase with exponential growth in the total number of deaths lasted approximately three weeks. To avoid the effect of low numbers, in this manuscript, the first day of this three week period of explosive growth is defined by the first day when the number of COVID-related deaths is 10 or higher. This approximately corresponds to the day when the community-spread phase of the epidemic starts. In the United States, the first COVID-19 related death goes back to a California's death on February 9, 2020, but the first day with 10 or more deaths was March 4 2020 (11 deaths); on March 25 2020, the total COVID-19 related death tall reached $1,260$ deaths. Thus, $1 \leq t \leq 21$ in Equation (1). This three week phase corresponds to approximately exponential growth in the total numbers of deaths. At the same, time there are countries where gradual change in death rate started right from the beginning of the three week change, see USA and Brazil COVID-19 total numbers of deaths on logarithmic-type scale in Figure 1.

[Figure 1 about here.]

The ordinary least squares (OLS) estimate, $\hat{\delta}_i = \left(\sum_{i=1}^{T} \sum_{t'=2}^{T} Y_{i,t} Y_{i,t'}\right) / \left(\sum_{i=1}^{T-1} Y_{i,t}^2\right)$, coincides with the maximum likelihood estimator of a simple linear regression model with normal errors, where $T$ is the total number of observations (21 days in the COVID-19 example). However, in contrast with the simple linear regression the (standardized) distribution of $\hat{\delta}_i$ does not converge to a normal distribution anymore due to auto-correlation. In White (1958), the limiting distribution of $\hat{\delta}_i$ was described via

$$\frac{\delta_i^T}{1+\delta_i^2}\left(\hat{\delta}_i - \delta_i\right) \xrightarrow{d} C(0,1), \tag{2}$$

when $\delta_i > 1$ (the outbreak phase of epidemic); $C(0,1)$ is the standard Cauchy random variable. In Phillips, Wu, and Yu (2009), this fact was used to build Cauchy-based confidence intervals for $\hat{\delta}_i$ in the analysis of NASDAQ bubble in the 1990s. With 21 days the distribution of $\hat{\delta}_i$ is still not Cauchy but clearly non-normal with heavy tails, see Figure 2.

[Figure 2 about here.]

What statistical tests work well for comparing the parameters of the explosive growth phase between different groups of countries? What test would be the best to use for comparing countries which enter into the pandemic later (LATE group) as compared to the countries who faced the pandemic earlier (EARLY group)?

### *1.2. Two-sample tests and asymptotic relative efficiency*

In most real-life situations, the underlying parametric distribution is not known and researchers cannot construct optimal likelihood ratio tests (LRT) or asymptotically optimal generalized likelihood ratio tests (GLRT). Permutation tests is a viable nonparametric solution. Permutations are used to evaluate approximate distributions of tests statistics under the null hypothesis, and consequently, secure the desired type 1 error, whereas power properties depend on a chosen test statistic Lehmann (1986).

When data are not normal, a widely popular alternative to two-sample t-test is Wilcoxon-Mann-Whitney (WMW) test Mann and Whitney (1947), which have been cited more ten thousand times since 1947. WMW test statisitc is a special case of U-statistics Hoeffding (1948).

Finite sample power properties of tests are often with Monte-Carlo simulation studies or calculated exactly if mathematics is not overly complex. Large sample power properties are described by an asymptotic relative efficiency (ARE) Nikitin (1995). We consider the ARE based on Pitman contiguous alternatives. WMW test secures $95\%$ asymptotic relative efficiency against the two-sample t-test under normal model Pitman (1949). ARE is not always that high: if two normal populations differ only in their variances, the power of WMW test is very low.

Asymptotic normality of U-statistics and its variations is traditionally used to find its asymptotic relative efficiency and asymptotic power Weinberg and Lagakos (2000), Rempala and Looney (2006), Chung, Romano, et al. (2013), Chung and Romano (2016). Nikitin (1995) (Theorem 1.4.1, page 15) defined Pitman ARE only for asymptotically normal test statistics.

Surprisingly, as shown in Section 8.2 some asymptotically normal U-statistics have asymptotic relative efficiency equal to zero, which means that asymptotic power under contiguous alternatives converges to $\alpha$ (type I error). At the same time, as simulation studies show (see Section 2), permutation tests based on these U-statistics have power properties different from $\alpha$. Why? The reason is that there are multiple test statistics with different local asymptotic properties leading to the same permutation test. Details are in Section 4.

Aslan and Zech (2005) proposed the *energy test* and used extensive Monte-Carlo experiments to assess its power properties. The test showed good statistical power properties when one of the populations had a Cauchy distribution, which makes it particularly interesting for explosive auto-regressive processes. When data are asymptotically normal, permutation tests result in high power properties. However, the application of U-test statistics for comparing *heavy-tailed* distributions often results in low statistical power. To extend the application of U-statistics, this manuscript explores the use of U-statistics with logarithm-based kernels, variants of energy tests proposed in Aslan and Zech (2005). As it will be shown in Section 2.4, use of such test statistics for distributions with Paretian tails does not change asymptotic normality test statistics. Interestingly, derivation of Pitman asymptotic relative efficiency cannot be deduced from asymptotic normality of these test statistics. In addition, power properties of these tests are also investigated through various Monte-Carlo simulation studies reported in Section 2. The manuscript conclude with a short summary in Section 7.

## 2. Two Sample Problem

Let $\mathbf{X} = (X_1, \ldots, X_{n_x})$ and $\mathbf{Y} = \left(Y_1, \ldots, Y_{n_y}\right)$ be two mutually independent samples of independent and identically distributed random variables with $X_i \stackrel{d}{=} F_X$ and $Y_j \stackrel{d}{=} F_Y$. Consider testing $H_0 : F_X = F_Y$ versus a composite alternative $H_A : F_X \neq F_Y$, where both

$F_X$ and $F_Y$ are absolutely continuous, so that their densities exist. It is convenient consider a pooled sample,

$$\mathbf{D} = (\mathbf{X}, \mathbf{Y}) = \left(X_1, \ldots, X_{n_x}, Y_1, \ldots, Y_{n_y}\right) = \left(D_1, \ldots, D_{n_x}, D_{n_x+1}, \ldots, D_n\right).$$

Realizations of random variables $X_i$, $Y_j$, $Z_{ij}$, and $D_i$ will be denoted by $x_i$, $y_j$, $z_{ij}$, and $d_i$, respectively.

This manuscript focuses on $X_i$ and $Y_j$ from the location-scale family, where $\mu_x$ and $\mu_y$ are location parameters and $\sigma_x$ and $\sigma_y$ are scale parameters of $F_X$ and $F_Y$, respectively. Then, the null $H_0 : X \sim f(d|\mu, \sigma), Y \sim f(d|\mu, \sigma)$ is tested versus the alternative $H_A : X \sim f(d|\mu_x, \sigma_x), Y \sim f(d|\mu_y, \sigma_y)$.

### *2.1. A class of test statistics, defined by the operator $T(g)$*

Consider a class of two-sample $U$-statistics defined by a kernel $g(Z_{ij})$,

$$T_{2u}(g) = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right), \tag{3}$$

where $Z_{ij}$ is a distance between $X_i$ and $Y_{j-n_x}$. For example, $Z_{ij} = |X_i - Y_{j-n_x}|$, $i = 1, \ldots, n_x$, $j = n_x + 1, \ldots, n$, $n = n_x + n_y$. The $g(|x_i - y_j|)$ can be interpreted as a distance between points $x_i$ and $y_j$. Thus, $T_{2u}(g)$ is an averaged distance.

The finite sample distribution of $T_{2u}(g)$ can very complex, and the permutation test represents a viable alternative to analytic derivation of such finite sample distributions. Asymptotic mean and variance of $T_{2u}(g)$ under continuous alternatives are found in Section 8.2.

$T_{2u}(g)$ is related by the energy test statistics, defined in Aslan and Zech (2005) as

$$T_{en}(g) = \frac{1}{n_x^2} \sum_{i=1}^{n_x} \sum_{j=i+1}^{n_x} g\left(|X_i - X_j|\right) + \frac{1}{n_y^2} \sum_{i=1}^{n_y} \sum_{j=i+1}^{n_y} g\left(|Y_i - Y_j|\right) - \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right), \tag{4}$$

but $T_{2u}(g)$ only uses its last term. In addition, $T_{en}(g)$ is also different from a regular one-sample U-statistic with kernel $g(Z_{ij})$:

$$\begin{aligned} T_{1u}(g) &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} g\left(|D_i - D_j|\right) = \frac{2}{n_x(n_x-1)} \sum_{i=1}^{n_x} \sum_{j=i+1}^{n_x} g\left(|X_i - X_j|\right) \\ &+ \frac{2}{n_y(n_y-1)} \sum_{i=1}^{n_y} \sum_{j=i+1}^{n_y} g\left(|Y_i - Y_j|\right) + \frac{2}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right), \end{aligned} \tag{5}$$

When the class of test statistics, $T_{en}(g)$ was suggested in Aslan and Zech (2005) the emphasis was made on the special case of $g\left(z_{ij}\right) = \log\left(|z_{ij}|\right)$, which is currently known as the energy test and is popular in applied literature. To eliminate the possibility of high negative values, authors also suggested an extension: $g\left(z_{ij}\right) = \log\left(\tau + |z_{ij}|\right)$, where $\tau$ is a small positive value. They showed that this test works well for many distributions including the standard Cauchy distribution. We build on this observation and focus on a more general class of heavy-tailed distributions: distributions with Paretian tails. The two-sample U-statistics

$$L_\gamma = \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} \ln(1 + |d_i - d_j|^\gamma),$$

also belongs to $T_{2u}(g)$ with $g(z) = \ln(1+|z|^\gamma)$. In our manuscript we mostly focus on $\gamma = 2$. Interestingly, the $WMW = \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} I(d_i - d_j > 0)$ test statistic does not belong to $T_{2u}(g)$ because $g(|d_i - d_j|)$ cannot be expressed via $I(d_i < d_j)$.

### *2.2. Likelihood Ratio Test (LRT)*

By Neyman-Pearson lemma the LRT statistic (negative two log-likelihood)

$$l = -2 \sum_{i=1}^{n_x} \log \frac{f(d_i|\mu, \sigma)}{f(d_i|\mu_x, \sigma_x)} + 2 \sum_{i=n_x+1}^{n} \log \frac{f(d_i|\mu, \sigma)}{f(d_i|\mu_y, \sigma_y)}$$

generates the most powerful test for testing a simple null versus a simple alternative .

The LRT for Cauchy distribution with $n_x = n_y$, $\sigma = \sigma_x = \sigma_y = 1$ and $\mu = \mu_y = 0$

$$l\left(\mu_x\right) = -2 \sum_{i=1}^{n_x} \log \frac{\left(d_i - \mu_x\right)^2 + 1}{d_i^2 + 1}. \tag{6}$$

For illustrative purposes, we consider a Cauchy example with $n_x = n_y = 50$ and $h = 4$, where $h = \mu_x \sqrt{n_x}$. This example will be denoted as **a simple Cauchy example**.

### *2.3. Generalized Likelihood Ratio Test (GLRT)*

Let $\hat{\mu}_x$ and $\hat{\sigma}_x$ [$\hat{\mu}_y$ and $\hat{\sigma}_y$] be the maximum likelihood estimators (MLEs) of $\mu_x$ and $\sigma_x$ [$\mu_y$ and $\sigma_y$], respectively, estimated on $\mathbf{X}$ [$\mathbf{Y}$] under $H_A$. Under $H_0$, the $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs of $\mu = \mu_x = \mu_y$ and $\sigma = \sigma_x = \sigma_y$ estimated on the pooled sample $\mathbf{D}$.

The GLRT statistic becomes

$$\hat{l} = -2 \sum_{i=1}^{n_x} \log \frac{f\left(d_i|\hat{\mu}, \hat{\sigma}\right)}{f\left(d_i|\hat{\mu}_x, \hat{\sigma}_x\right)} - 2 \sum_{i=n_x+1}^{n} \log \frac{f\left(d_i|\hat{\mu}, \hat{\sigma}\right)}{f\left(d_i|\hat{\mu}_y, \hat{\sigma}_y\right)}. \tag{7}$$

When permutations are applied to evaluate the distribution of $\hat{l}$ under $H_0$, the values of $\hat{\mu}$ and $\hat{\sigma}^2$ do not change, making $\sum_{i=1}^{n_x} \log f(d_i|\hat{\mu}, \hat{\sigma})$ and $\sum_{i=n_x+1}^{n} \log f(d_i|\hat{\mu}, \hat{\sigma})$ invariant to permutations. Thus, the test statistics $\hat{l}$ is simplified to

$$\hat{l}^* = \sum_{i=1}^{n_x} \log f\left(d_i|\hat{\mu}_x, \hat{\sigma}_x\right) + \sum_{i=n_x+1}^{n} \log f\left(d_i|\hat{\mu}_y, \hat{\sigma}_y\right). \tag{8}$$

The **GLRT for Cauchy distributions** is

$$\hat{l} = -2\sum_{i=1}^{n_x} \log \frac{\hat{\sigma}_x \pi^{-1}}{(d_i - \hat{\mu}_x)^2 + \hat{\sigma}_x^2} - 2\sum_{i=n_x+1}^{n} \log \frac{\hat{\sigma}_y \pi^{-1}}{(d_i - \hat{\mu}_y)^2 + \hat{\sigma}_y^2} + 2\sum_{i=1}^{n} \log \frac{\hat{\sigma} \pi^{-1}}{(d_i - \hat{\mu})^2 + \hat{\sigma}^2}. \tag{9}$$

The MLEs are found numerically using box constrained maximization quasi-Newton optimization Zhu, Byrd, Lu, and Nocedal (1995). When permutation tests are considered the generalized log-likelihood ratio test is simplified to

$$\hat{l}^* = -\frac{1}{n}\sum_{i=1}^{n_x} \log \frac{\hat{\sigma}_x}{(d_i - \hat{\mu}_x)^2 + \hat{\sigma}_x^2} - \frac{1}{n}\sum_{i=n_x+1}^{n} \log \frac{\hat{\sigma}_y}{(d_i - \hat{\mu}_y)^2 + \hat{\sigma}_y^2} \tag{10}$$

### *2.4. Heavy tailed distributions (Paretian tails)*

If $X_i$ and $Y_j$ belong to the family of stable distributions (Normal, Cauchy, Levy, etc.) their second moments do not exist if the stability parameter $< 2$ (non-normal distributions). However, for some choices of $g(\cdot)$ all moments of $g(Z_{ij})$ do exist and asymptotic normality holds. For example, for stable $Z_{ij}$ with $g(Z_{ij}) = \log\left(1 + Z_{ij}^{\gamma}\right)$, $\gamma > 0$, moments exist as follows from Theorem 2.1. Hence, asymptotic normality of $T_{2u}(g)$ holds and researchers can apply a normal approximation to a distribution of $T_{2u}(g)$ when sample size becomes large and permutation tests become cost prohibitive.

For an absolutely continuous random variable $Z$ to have *Pareto tails*, we require $\forall \epsilon > 0$ there exists $M > 0$ so that $\forall x > M$,

$$\left| \frac{\partial}{\partial x} \text{Pr}\left(|Z| > x\right) - \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}} \right| < \epsilon$$

with some shape parameter $\alpha$, also called *a tail index*, and a scale parameter $x_m$.

Stable distributions only have moments up to order $\alpha$, but all logarithmic moments do exist, see Uchaikin and Zolotarev (2011). Theorem 2.1 considers a more general class of distributions: distributions with Paretian tails. Another difference is the use of $\gamma$ power.

**Theorem 2.1.** *If $Z$ is a random variable with Paretian tails, then $\forall \gamma, p > 0$, $E\left[\log\left(1 + |Z|^{\gamma}\right)\right]^p < \infty$.*

Proof of Theorem 2.1 is moved to the Appendix. Theorem 2.1 easily extends to $\text{E}\left[\log\left(\tau + |Z|^{\gamma}\right)\right]^p < \infty$ for any positive finite $\tau$.

## 3. Asymptotic efficiency for testing location shift for Cauchy data: non-permuted tests

### *3.1. Likelihood Ratio Test*

Le Cam (1960) showed local asymptotic normality of twice differentiable log-likelihood under contiguous alternatives. Under $H_A : \mu_x = h/\sqrt{n_x}$ with $h > 0$,

$$l\left(\mu_x\right)\Big|_{H_A} \xrightarrow{d} -2 \cdot N\left(-\frac{1}{2}h^2 I\left(0\right), h^2 I\left(0\right)\right).$$

For Cauchy's distribution with unit scale Fisher's information

$$I(0) = I(\mu_x) = \int_{-\infty}^{\infty} 4x^2\pi^{-1}(1+x^2)^{-3}dx = 0.5,$$

$l(\mu_x) \xrightarrow{d} N\left(-h^2/2, 2h^2\right)$ and the asymptotic power is

$$\Pr\left(\frac{l(\mu_x)+h^2/2}{h\sqrt{2}} > z_{1-\alpha}\middle|\mu_x = h/\sqrt{n_x}\right) \to 1-\Phi\left(z_{1-\alpha}-\frac{h}{\sqrt{2}}\right). \quad (11)$$

By construction the efficiency of the $LRT$ test is Eff$(LRT) = 1/\sqrt{2}$, when $h = \mu_x\sqrt{n_x}$. If another definition of the local alternative is used: $h = \mu_x\sqrt{n} = \mu_x\sqrt{n_x}\sqrt{n/n_x}$, then Eff$(LRT) = 1/\sqrt{2/p_x}$.

**The simple Cauchy example:** if $h = 4$, $n_x = n_y$, then for one sided testing (type 1 error $=\alpha$) of $H_0 : X, Y \sim C(0,1)$ versus $H_0 : X \sim C(0,1), Y \sim C(h/\sqrt{n}, 1)$ the asymptotic power $1-\Phi\left(z_{1-\alpha} - 4/\sqrt{2}\right) = 88.2\%$ is highest by Nayman-Pearson lemma. Then, ARE$(LRT) = 1$.

### *3.2. Generalized Likelihood Ratio Test*

When the MLEs $\hat{\mu}_x$ is used instead of the unknown $\mu_x$, the asymptotic distributions of $n^{-1/2}l(\mu_x)$ and $n^{-1/2}l(\hat{\mu}_x)$ are asymptotically equivalent. This directly follows from consistency and asymptotic normality of Cauchy MLEs and the expansion

$$l(\mu_x) \quad = \quad l(\hat{\mu}_x) + l'(\hat{\mu}_x)(\hat{\mu}_x - \mu_x) + l''(\hat{\mu}_x)\frac{(\hat{\mu}_x-\mu_x)^2}{2} + o_P\left[(\hat{\mu}_x-\mu_x)^2\right]. \quad (12)$$

which means that Eff$(GLRT) =$ Eff$(LRT)$.

**The simple Cauchy example:** If $p_x = p_y = 0.5$, $\sqrt{n_x}\,(\hat{\mu}_x - \mu_x) \xrightarrow{d} N(0,2)$ and Eff$(GLRT) =$ Eff$(LRT) = 1/\sqrt{2}$. Then, ARE$(GLRT) = 1$.

### *3.3. Asymptotic efficiency of $L_2$ at $\mu_x = h/\sqrt{n_x}$*

As shown in Appendix (8.3) and (8.4) with $\mu_x = \frac{h}{\sqrt{n_x}}$ and $\mu_y = 0$,

$$E_{\mu_x}(L_2) = \log(9) + \frac{2}{9}\frac{h^2}{n} + O\left(n^{-3/2}\right)$$

and

$$\sigma^2_{\mu_x}(L_2) := \lim_{n\to\infty} n \cdot Var(L_2) = p_x^{-1}C_x + O\left(n^{-1}\right),$$

where $C_x \approx 2.053755$.

**The simple Cauchy example:** From $E'_{\mu_x}(L_2) = 2h/(9n_x) + O\left(n^{-3/2}\right)$, we find $E'_{\mu_x=0}(L_2) = O\left(n^{-1}\right)$. Then, Eff$(L_2) =$ ARE$(L_2) = 0$.

### 3.4. Asymptotic efficiency of $T_{2u}(g)$ at $\mu_x = h/\sqrt{n_x}$

As shown in Appendix (8.5) and (8.6) under $\mu_x = \frac{h}{\sqrt{n_x}}$ and $\mu_y = 0$,

$$E_{\mu_x}(T_{2u}(g)) = c_1 + c_2 \frac{h_x^2}{2n} + O\left(n^{-3/2}\right)$$

and

$$\sigma^2_{\mu_x}(T_{2u}(g)) := \lim_{n\to\infty} n \cdot Var(L_2) = c_4 + O\left(n^{-1}\right),$$

where $c_1$, $c_2$ and $c_4$ are some constants.

**The simple Cauchy example:** From $E'_{\mu_x}(T_{2u}(g)) = c_2 h_x/n + O\left(n^{-3/2}\right)$, we find $E'_{\mu_x=0}(T_{2u}(g)) = O\left(n^{-1}\right)$. Then, Eff$(T_{2u}(g)) =$ ARE$(T_{2u}(g)) = 0$.

## 4. Asymptotic efficiency of permutation tests

It is important to mentions that the families of tests, $T_{en}$, $T_{1u}$, and $T_{2u}$ are permutation equivalent and share the same power properties. Consequently, its special cases with $g(|d_i - d_j|) = \log(1 + |d_i - d_j|^2)$ (in this case $L_2 = T$) also share the same asymptotic properties.

### 4.1. Asymptotic efficiency of "permuted" $L_2$ at $\mu_x = h/\sqrt{n_x}$

Power properties of the "permuted test" $\{L_2 > c^*(L_2)\}$ are different from its non-permuted version $\{L_2 > c\}$, because for permuted test $c^*(L_2)$ is a random variable: $c^*(L_2) = G_n^{-1}(1 - \alpha|L_2)$, where $G_n$ is a permuted distribution of $L_2$ under $H_0$; $G_n^{-1}$ is its inverse. Section 5 shows that ARE of permuted $L_2$ is not zero for location families and is comparable to WMW test. Test $\{L_2 > c^*(L_2)\}$ is equivalent to $\{G_n(L_2) > 1 - \alpha\}$. Under $H_0$, $G_n(L_2)$ is a uniformly distributed (discrete) random variable (by the property of permuted distributions), which means that $\Phi^{-1}(1 - G_n(L_2))$ is a standard normal random variable under $H_0$, and the test $\left\{\Phi^{-1}(1 - G_n(L_2)) > \Phi^{-1}(1 - \alpha)\right\}$ is equivalent to permuted $\{L_2 > c^*(L_2)\}$.

### 4.2. P-value invariant "permuted" tests

The transformation $G_n(\cdot)$ generates P-values of permutation tests. The permutation tests based on test statistics $T_{en}$, $T_{1u}$ and $T_{2u}$ share the same power properties

$$\Pr(T_{en} > c^*(T_{en})|\theta) = \Pr(T_{1u} > c^*(T_{1u})|\theta) = \Pr(T2u > c^*(T2u)|\theta)$$

because P-values of the permutation tests are the same: $G_n(T_{en}) = G_n(T_{1u}) = G_n(T_{2u})$.

[Figure 3 about here.]

In Figure 3, Panels (a) and (b) show histograms of distributions of $L_2$ and permutation critical values. Both are approximately normally distributed, but they are also highly corrected, as shown in Panel (c). Panel (d) shows the distribution of permutation P-values, which is very different from uniform one would expect under $H_0$; dashed lines trace significant level of $0.05$ and observed power of the permutation test $0.356$. Distributions of P-values indicate that asymptotic relative efficiency is different from 0. Since $G_n$ converges to a standard uniform random variable on unit interval, an inverse standard normal transformation ensures that $\Phi^{-1}(G_n) \xrightarrow{d} N(0,1)$ under $H_0$. The estimated cumulative distribution function of $\Phi^{-1}(G_n)$ under $h = 4$ and $n_x = 80$ is presented at Panel (e). Vertical dashed line is drawn at a critical value (= 1.96), horizontal lines dashed lines show the observed power and the power estimated under a normal model.

**Definition:** Test statistics with the same permutation P-values are called permutation test invariant.

For example, $T_{en}(g)$, $T_{1u}(g)$ and $T_{2u}(g)$ are permutation test invariant as they share exactly the same permutation P-values. Similarly, $l$ and $l^*$ is another example of permutation test invariant test statistics.

**Theorem 4.1.** *If $\Phi^{-1}(G_n)$ is asymptotically normal random variable with a continuous and differentiable in the neighbourhood of $\theta = 0$ asymptotic mean then Pitman's efficiency (slope) for the whole group of permutation invariant test statistics is*

$$\text{Eff}(\Phi^{-1}(G)) = \lim_{n\to\infty} E'_{h/\sqrt{n}}\left(\Phi^{-1}(G_n)\right).$$

**Proof:** The group of permutation invariant test statistics associated with the same permutation test includes the test statistic $\Phi^{-1}(G_n)$. The strictly monotone transformation $\Phi^{-1}(\cdot)$ does not depend on $\theta$ and continues to be the same transformation under the null and alternative hypotheses. The transformation $\Phi^{-1}(\cdot)$ maps random variables $G_n$ (p-values) into the standard normal random variable under $H_0$. Generally, under the alternative the asymptotic distribution of $\Phi^{-1}(\cdot)$ maybe different from normal, which is why the theorem requires asymptotic normality of $\Phi^{-1}(\cdot)$.

Thus, $\Phi^{-1}(\cdot)$ extracts an asymptotically normal permutation invariant test statistics with a standard normal distribution under $H_0$ with a differentiable mean in the neighbourhood of $\theta$.

The *efficiency* is defined as

$$\text{Eff}(\Phi^{-1}(G_n)) = \frac{E'_{\theta=0}\left(\Phi^{-1}(G_n)\right)}{\sigma_{\theta=0}(\Phi^{-1}(G_n))}, \tag{13}$$

where the subscript $\theta = 0$ refers to a limiting quantity at $n = \infty$. After $\Phi^{-1}(\cdot)$ transformation $\sigma_{\theta=0}(\Phi^{-1}(G_n)) = 1$. **Q.E.D.**

The $\text{Eff}(\Phi^{-1}(G))$ can be approximated by $h^{-1}\hat{E}_{h/\sqrt{n}}\left(\Phi^{-1}(G_n)\right)$, where $\hat{E}_{h/\sqrt{n}}$ is an approximated value of $E_{h/\sqrt{n}}$.

**Simple Cauchy Example:** under $h = 4$, the $\text{Eff}(\Phi^{-1}(G)) \approx 0.273$ based on a Monte-Carlo study with 1000 repetitions and $n = 80$. Then, the asymptotic power of the permutation test is $\approx 1 - \Phi(z_{1-\alpha} - 4 \cdot 0.273) \approx 0.290$. This approximation is somewhat different from what a simulation study showed (power = $0.356$) but it relies on large sample approximation.

## 5. Power comparisons: Monte-Carlo simulations

In Table 1, Monte-Carlo studies explore convergence to asymptotic power under local alternatives in location parameter of Cauchy distributions.

All experimental settings used $n_x = n_y = 50$ and type 1 error of $5\%$.

[Table 1 about here.]

Table 2 reports results of Monte-Carlo simulations when data are generated from three stable distributions (Normal $N(\mu, \sigma)$, Cauchy $C(\mu, \sigma)$, and Levy $L(\mu, \sigma)$) and from a distribution with super-heavy tails (log-Cauchy $LC(\mu, \sigma)$).

[Table 2 about here.]

Each Monte-Carlo experiment used $m = 2500$ resamples to evaluate power properties, leading to the Monte-Carlo error $\hat{\sigma}_m = \sqrt{\hat{p}(1-\hat{p})/m}$ and Wald's asymptotic $95\%$ confidence intervals $(\hat{p} - 2\hat{\sigma}_m,\, \hat{p} + 2\hat{\sigma}_m)$, where $\hat{p}$ is a Monte-Carlo power estimate. In a most conservative case of $p = 0.5$, $\hat{\sigma}_m = 0.01$ and the confidence interval is approximately equal to $(\hat{p} - 0.02,\, \hat{p} + 0.02)$.

Since the number of permutations quickly grows with sample size, at $n_x > 20$ or $n_y > 20$, we will rely on randomized permutation test with 1600 resamples. This number was suggested by Keller-Mcnulty and Higgins (1987) and used by Marozzi (2004)).

Table 2 reports simulation results from three stable distributions (Normal, Cauchy and Levy) and log-Cauchy. In this table one sample is generated from the distribution $F_1$ and another is from $F_2$. Power properties of test staistics $L_\gamma$ ($\gamma \in \{0.5, 1, 2\}$) are compared with power properties of likelihood ratio tests built under Normal, Cauchy and Levy distributional assumptions. Wilcoxon-Mann-Whitney and Kolmogorov-Smirnov tests are added as popular alternative solutions for two group comparisons.

The $lrt_{distribution}$ statistics assumed that both location and scale parameters can differ between populations, this is why occasionally Wilcoxon-Mann-Whitney outperformed $L_{norm}$, which initially may seem counter-intuitive as it is well known that asymptotic relative efficiency of Wilcoxon-Mann-Whitney test is $95.5\%$. For example, when comparing $N(0, 1)$ versus $N(0.6, 1)$ power of Wilcoxon-Mann-Whitney test was $82\%$, whereas $lrt_n$ only showed $75\%$ power. However, when it is assumed that scales (variances) between the two normal populations are the same and the difference can only happen between their locations (means), the LRT test is different from $lrt_n$ and is equivalent to two sample t-test securing $84\%$ power. Under even stronger assumptions that variances are the same and equal to 1, z-test can be used with even higher power of $86\%$.

Log-Cauchy is another heavy tailed distribution, but it has heavier tails than the family of stable distributions, sometimes referred as "super-heavy". The logarithmic kernel in $L_\gamma$ does not fully suppress the impact of heavy tails. The tails are heavier than Pareto tails and Theorem 2.1, which justifies asymptotic normality of $L_\gamma$, is not applicable any longer.

## 6. Comparison of explosive phase parameters between countries with EARLY and LATE entries into the pandemic

To accommodate overtime change in death rate, $\delta$ can be assumed to be different for every country and change during this three week period. The parametarization

$$\delta_i = \beta_{i0} + \beta_{i1} t$$

is used to accommodate time effect, where $\beta_1 > 0$ describes overtime increase death rate and $\beta_1 < 0$ shows decrease. Thus, $\beta_1$ describes the efficiency of early response to COVID-19 epidemic.

[Figure 4 about here.]

Figure 4 shows that similarly to a one parameter in Figure 2, distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ are also non-normal and heavy tailed.

Now we expand the problem settings towards multiple countries enumerated by index $i$. For each country we extract numbers of deaths in the first 21 days as described above and apply the first order serial auto-correlation with $\delta_i = \beta_{0i} + \beta_{1i}t$ for each country. Distributions of $\hat{\beta}_{0i}$ (and $\hat{\beta}_{1i}$) are independent with finite variance.

[Figure 5 about here.]

The scatter-plot of $\hat{\beta}_{i0}$ and $\hat{\beta}_{i1}$ of 86 countries which experienced at least three weeks with daily COVID-related deaths of 10 or more is reported on Figure 5. Of these 86, 39 countries with the first day of 10 or more deaths prior to March 28 2020 (EARLY group) are plotted by solid circles and 47 countries with the first day of 10 or more COVID-related deaths on or after March 28 2020 (LATE group) is plotted by empty circles. The values $\hat{\beta}_{i0}$ and $\hat{\beta}_{i1}$, $(i = 1, \ldots, 86)$ are given in Tables 4 (EARLY outbreak) and 5 (LATE outbreak).

Both Cauchy-based likelihood ratio test and $L_2$ permutation test gave P-values $< 0.0001$, confirming different mortality patterns between countries entered epidemic earlier than later. For the countries in EARLY groups averaged and mean value estimates of $\beta_0$ and $\beta_1$, we $\hat{\beta}_0^{ave,early} = 1.2625$ and $\hat{\beta}_0^{med,early} = 1.2254$, $\hat{\beta}_0^{ave,late} = 1.1354$ and $\hat{\beta}_0^{med,late} = 1.1279$, respectively. Interpreting medians, prior to March 28 2020, on the first day of 10 or more COVID-19 related deaths, the countries experiences 22.52% daily increase in the total number of COVID-19-related deaths. Whereas on or after March 28 2020, the first day with 10 or more deaths was characterized by 12.79% daily increase in the number of deaths.

Percent daily increases in the total number of deaths decrease overtime with reductions characterized by, $\hat{\beta}_1^{ave,early} = -0.0080$, $\hat{\beta}_1^{med,early} = -0.0077$, $\hat{\beta}_1^{ave,late} = -0.0047$, and $\hat{\beta}_1^{med,late} = -0.0047$.

The *plateau* in the number of deaths is defined by $\delta = 1$. Using simple algebra we can estimate that for EARLY group the plateau is reached after approximately 29 days $(0.2254/0.0077 = 29.2727)$ after the first day of 10+ deaths. For the LATE group, the plateau is estimate to be reached on 27th day $(0.1279/0.0047 = 27.2128)$.

Overall, we conclude that countries which experienced a later onset of COVID-19 epidemic had a lower rate on increase in COVID-19 related deaths but the reduction of this increase was also lower as compared to countries with early onset of the COVID-19 epidemic.

## 7. Summary

While working on statistical modeling of the explosive phase of COVID-19 epidemic, we have applied auto-regression model to this initial epidemic phase. This model allowed us to estimate country-specific rates and their changes during the first three weeks of the COVID-19 epidemic.

The distribution of regression coefficients of auto-regression models during the explosive epidemic phase has heavy tails. Unknown distribution of residuals motivated us to explore non-parametric approaches to comparing groups of countries. Asymptotic convergence of regression coefficients to Cauchy distribution indicated that more attention needs paid to the Cauchy case.

To our surprise we found that Pitman's asymptotic efficiency of a the test statistic $[T_{2u}(g)]$ based on averaged distances between the elements of two samples is equal zero. This finding was in a strike contrast with excellent performance of the permutation test based on $T_{2u}(g)$ in various simulation studies. The power of $T_{2u}(g)$-based permutation test was comparable with the power of Wilcoxon-Mann-Whitney test for location shift alternatives and the power was substantially higher for detecting changes in scale.

We found that large sample normality of a test statistic is not sufficient to correctly quantify its asymptotic power properties. There are many test statistics leading to the same permutation test. The problem is that each of these test statistics has has it own asymptotic distribution. Some of these distributions are normal, some are not, some test statistics may diverge, but they still produce the same permutation test. Theorem 4.1 stated what test statistic should be used for determining asymptotic efficiency of a permutation test.

Overall, we urge readers to be careful while exploring asymptotic properties of permutation tests. Asymptotic normality of a test statistic maybe insufficient for correct assessment of its asymptotic efficiency, relative efficiency and power.

## 8. Appendix

### *8.1. Proof of Theorem 2.1*

To show that $\mathrm{E}\left[\log\left(1+|Z|^{\gamma}\right)\right]^{p}$ is finite we use

$$\mathrm{E}\left[\log\left(1+|Z|^{\gamma}\right)\right]^{p} \;=\; \mathbf{E}\left[I_{|Z|<M}\log\left(1+|Z|^{\gamma}\right)\right]^{p}+\mathrm{E}\left[I_{|Z|>M}\log\left(1+|Z|^{\gamma}\right)\right]^{p}. \quad (14)$$

Further we consider only upper tail with $Z > M$. The first term in (14) is finite, the second term $\mathrm{E}\left[I_{|Z|>M}\log\left(1+Z^{\gamma}\right)\right]^{p}$ is dominated above by

$$\begin{aligned}
&\int_{I_{|z|>M}}\left[\log\left(1+z^{\gamma}\right)\right]^{p}\frac{\alpha x_m^{\alpha}}{z^{\alpha+1}}dz+\epsilon \mathrm{Pr}\left(|z|>M\right)\\
=&\int_{I_{|z|>M}}\left[\frac{\log\left(1+z^{\gamma}\right)}{\log\left(z^{\gamma}\right)}\right]^{p}\left[\log\left(z^{\gamma}\right)\right]^{p}z^{-\alpha-1}\alpha x_m^{\alpha}dz+\epsilon \mathrm{Pr}\left(|z|>M\right)\\
=&\int_{I_{|z|>M}}\left[\frac{\log\left(1+z^{\gamma}\right)}{\log\left(z^{\gamma}\right)}\right]^{p}\gamma^{p}\left[\log\left(z\right)\right]^{p}z^{-\alpha-1}\alpha x_m^{\alpha}dz+\epsilon \mathrm{Pr}\left(|z|>M\right) \qquad (15)
\end{aligned}$$

Let $G_p(z)=\left[\log\left(z\right)\right]^{p}z^{-\alpha-1}$. Sequential application of L'Hospital's rule

$$\begin{aligned}
\lim_{z\to\infty}G_p(z) &= \lim_{z\to\infty}\frac{\frac{\partial}{\partial z}\left[\log\left(z\right)\right]^{p}}{\frac{\partial}{\partial z}z^{\alpha+1}}=\lim_{z\to\infty}\frac{p\left[\log\left(z\right)\right]^{p-1}z^{-1}}{\alpha z^{\alpha}}\\
&= \lim_{z\to\infty}\frac{p\left[\log\left(z\right)\right]^{p-1}}{\alpha z^{\alpha+1}}=\lim_{z\to\infty}\frac{p\left[\log\left(z\right)\right]^{p-1}}{\alpha z^{\alpha+1}}=\lim_{z\to\infty}\frac{p}{\alpha}G_{p-1}(z)
\end{aligned}$$

gives a recursive relationship which ensures that $G_p(z)$ converges down to zero as $z\to\infty$, as

$$\lim_{z\to\infty}G_p(z)=\lim_{z\to\infty}\frac{p^p}{\alpha^p}G_0(z)=\frac{p^p}{\alpha^p}\lim_{z\to\infty}\frac{1}{z^{\alpha+1}}=0.$$

Other terms in the integrand of the integral of (15) also converge either to zero or to a constant. Thus, for a sufficiently large $M$ there exists $\epsilon'$ so that

$$\left[\frac{\log\left(1+z^{\gamma}\right)}{\log\left(z^{\gamma}\right)}\right]^{p}<\epsilon'$$

and, going back to two tails with $Z > M$ (upper) and $Z < -M$ (lower),

$$\begin{aligned} \mathrm{E}\left[I_{|Z|>M} \log\left(1+Z^{\gamma}\right)\right]^{p} &< \int_{I_{|z|>M}} \left[\log\left(1+z^{\gamma}\right)\right]^{p} \frac{\alpha x_m^{\alpha}}{z^{\alpha+1}} dz + \epsilon \mathrm{Pr}\left(|z|>M\right) \\ &< \epsilon' \int_{I_{|z|>M}} \frac{\alpha x_m^{\alpha}}{z^{\alpha+1}} dz + \epsilon \mathrm{Pr}\left(|z|>M\right) < \epsilon' + \epsilon \end{aligned} \tag{16}$$

which ensures that the integral (15) is finite when $\alpha > 0$. **Q.E.D.**

### *8.2. Finite and large sample properties of $T_{2g}(g)$*

Let the first two moments of $g\left(Z_{ij}\right)$ exist and be finite, $Z_{ij} = |X_i - Y_j|$ and $g\left(\cdot\right)$ be a non-negative monotone function. Below, we derive, asymptotic mean and variance of $T = T_{2u}(g)$

[Table 3 about here.]

#### *8.2.1. Finite $n_x$ and $n_y$: independent observations*

If $Cov\left(g\left(Z_{i_1 j_1}\right), g\left(Z_{i_2 j_2}\right)\right)$ takes different values for different choices of $i_1, i_2 \in \{1, \cdots, n_x\}$ and $j_1, j_2 \in \{n_x+1, \cdots, n\}$, then for every finite $n_x$ and $n_y$,

$$ET = \frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} E\left(g\left(Z_{ij}\right)\right)$$

and

$$\begin{aligned} Var\left(T\right) = \frac{1}{n_x^2 n_y^2} Var\left(\sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right)\right) &= \frac{1}{n_x^2 n_y^2} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} Var\left(g\left(Z_{ij}\right)\right) \\ &+ \frac{1}{n_x^2 n_y^2} \sum_{i=1}^{n_x} \sum_{j_1=n_x+1}^{n} \sum_{j_2=n_x+1, j_1 \neq j_2}^{n} Cov\left(g\left(Z_{ij_1}\right), g\left(Z_{ij_2}\right)\right) \\ &+ \frac{1}{n_x^2 n_y^2} \sum_{i_1=1}^{n_x} \sum_{i_2=1, i_1 \neq i_2}^{n_x} \sum_{j=n_x+1}^{n} Cov\left(g\left(Z_{i_1 j}\right), g\left(Z_{i_2 j}\right)\right). \end{aligned}$$

#### *8.2.2. Finite $n_x$ and $n_y$: independent and identically distribution observations*

If $X_i$ are i.i.d. and $Y_j$ are i.i.d., $E\left(g\left(Z_{ij}\right)\right) = E_{(iid)}$ is the same for all combinations of $i$ and $j$, with

$$ET = E\left(\frac{1}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=n_x+1}^{n} g\left(Z_{ij}\right)\right) = E_{(iid)} \tag{17}$$

and

$$
\begin{aligned}
Var\,(T) &= V\frac{n_x n_y}{n_x^2 n_y^2} + C_x\frac{n_x n_y(n_y-1)}{n_x^2 n_y^2} + C_y\frac{n_y n_x(n_x-1)}{n_x^2 n_y^2} \\
&= V\frac{1}{n_x n_y} + C_x\frac{n_y-1}{n_x n_y} + C_y\frac{n_x-1}{n_x n_y}. \qquad (18)
\end{aligned}
$$

*8.2.3. Large sample properties*

Let $X_i$ be i.i.d., $Y_j$ be i.i.d., $p_x = \lim_{n\to\infty} n_x/n$, and $p_y = \lim_{n\to\infty} n_y/n$.

**Asymptotic properties of $\sqrt{n_x}T$ and $\sqrt{n_y}T$** If both $n_x$ and $n_y$ go to $\infty$, but $p_x = 0$ ($n_x$ goes to infinity slower than $n_y$), it is reasonable to consider a limiting distribution of $\sqrt{n_x}T$ instead of $\sqrt{n}T$. Then, as follows form Equations (17) and (18),

$$\sqrt{n_x}\,(T - ET) \xrightarrow{d} N(0, C_x).$$

Similarly, if $p_y = 0$,

$$\sqrt{n_y}\,(T - ET) \xrightarrow{d} N(0, C_y).$$

**Asymptotic properties of $\sqrt{n}T$** Asymptotic properties of $\sqrt{n}T$ can be considered for three distinct cases: (1) $n_y$ is fixed while $n \to \infty$, (2) $n_x$ is fixed while $n \to \infty$, and (3) $n_x \to \infty$ and $n_y \to \infty$ with $0 < p_x < 1$. The first two cases are of lesser interest because $nVar(T)$ diverges. The third case, $0 < p_x < 1$, also directly follows from Equations (17) and (18)

$$\sqrt{n}\,(T - ET) \xrightarrow{d} N(0, p_x^{-1}C_x + p_y^{-1}C_y).$$

### 8.3. Mean of $L_2$ (simple Cauchy example)

In the case when $g(Z_{ij}) = \log\left(1 + Z_{ij}^2\right)$, the use of Taylor expansion leads to:

$$
\begin{aligned}
ET &= E\left(\log\left(1+Z_{ij}^2\right)\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \log\left(1+(x-y)^2\right) \\
&\times \left[\frac{1}{x^2+1} + \frac{2x}{(x^2+1)^2}\mu_x + \frac{6x^2-2}{(x^2+1)^3}\frac{\mu_x^2}{2} + O\left(\mu_x^3\right)\right] \\
&\times \left[\frac{1}{y^2+1}\right]\frac{dx}{\pi}\frac{dy}{\pi} \qquad (19)
\end{aligned}
$$

Using

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \log\left(1+(x-y)^2\right)\frac{1}{x^2+1}\frac{1}{y^2+1}dxdy = \log(9)\pi^2 \approx 21.6857, \quad (20)$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \log\left(1+(x-y)^2\right)\frac{1}{y^2+1}\frac{2x}{(x^2+1)^2}dxdy = 0, \qquad (21)$$

and

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\log\left(1+(x-y)^2\right)\frac{1}{\pi(y^2+1)}\frac{6x^2-2}{\pi(x^2+1)^3}dxdy=\frac{2}{9}. \tag{22}$$

Hence, $ET=\log(9)+\frac{1}{9}\mu_x^2+O\left(\mu_x^3\right)$. If $\mu_x=h/\sqrt{n}$, $ET=\log(9)+\frac{h^2}{9n}+O\left(n^{-3/2}\right)$.

### *8.4. Variance of $L_2$ (simple Cauchy example)*

$C_x=Cov\left(g\left(Z_{i_1j_1}\right),g\left(Z_{i_1j_2}\right)\right)=Cov\left(g\left(Z_{i_1j_1}\right),g\left(Z_{i_2j_1}\right)\right)=C_y,$

$$\lim_{n\to\infty}nVar\left(T\right) \;=\; (p_x^{-1}+p_y^{-1})C_x.$$

$$\begin{aligned}
C_y &= Cov\left[\log\left(1+Z_{ij}^2\right),\log\left(1+Z_{kj}^2\right)\right]\\
&= E\left[\log\left(1+Z_{ij}^2\right)\cdot\log\left(1+Z_{kj}^2\right)\right]-(ET)^2+O(n^{-2})\\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\log\left(1+(x_1-y)^2\right)\log\left(1+(x_2-y)^2\right)\\
&\times \left[\frac{1}{x_1^2+1}+\frac{2x_1}{\left(x_1^2+1\right)^2}\frac{h_x}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\left[\frac{1}{x_2^2+1}+\frac{2x_2}{\left(x_2^2+1\right)^2}\frac{h_x}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\\
&\times \left[\frac{1}{y^2+1}+\frac{2y}{\left(y^2+1\right)^2}\frac{h_y}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\frac{dx_1}{\pi}\frac{dx_2}{\pi}\frac{dy}{\pi}-(ET)^2+O(n^{-2})\\
&= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty}\log\left(1+(x-y)^2\right)\frac{1}{x^2+1}\frac{dx}{\pi}\right]^2\left[\frac{1}{y^2+1}+\frac{2y}{\left(y^2+1\right)^2}\frac{h_y}{\sqrt{n}}\right]\frac{dy}{\pi}\\
&+ \frac{4h_x}{\sqrt{n}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\log\left(1+(x_1-y)^2\right)\log\left(1+(x_2-y)^2\right)\\
&\times \left[\frac{1}{x_1^2+1}\right]\left[\frac{2x_2}{\left(x_2^2+1\right)^2}\right]\left[\frac{1}{y^2+1}+\frac{2y}{\left(y^2+1\right)^2}\frac{h_y}{\sqrt{n}}\right]\frac{dx_1}{\pi}\frac{dx_2}{\pi}\frac{dy}{\pi}\\
&- (ET)^2+O\left(\frac{1}{n}\right)
\end{aligned} \tag{23}$$

Note, the second summand in (23) is equal to zero, because

$$v(y)=\log\left(1+(y-z)^2\right)\left[\frac{2y}{\left(y^2+1\right)^2}\right]$$

is symmetric and $\int_0^\infty v(y)dy=-\int_{-\infty}^0 v(y)dy$. Then, using

$$\int_{-\infty}^{+\infty}\log\left(1+(x-z)^2\right)\frac{1}{x^2+1}\frac{dx}{\pi}=\log(4+z^2)$$

equation (23) simplifies to

$$C_y = \int_{-\infty}^{+\infty} \left[\log(4+z^2)\right]^2 \left[\frac{1}{z^2+1} + \frac{2z}{(z^2+1)^2}\frac{h_y}{\sqrt{n}}\right]\frac{dz}{\pi} - (ET)^2 + O\left(\frac{1}{n}\right) \quad (24)$$

Numeric integration shows that

$$\int_{-\infty}^{+\infty} \left[\log(4+z^2)\right]^2 \left[\frac{1}{z^2+1}\right]\frac{dz}{\pi} \approx 6.881531, \quad (25)$$

$(ET)^2 \approx 4.827776$, and due to symmetry arguments,

$$\int_{-\infty}^{+\infty} \left[\log(4+z^2)\right]^2 \left[\frac{2z}{(z^2+1)^2}\right]\frac{dz}{\pi} = 0 \quad (26)$$

leading to

$$C_y = 6.881531 - 4.827776 + O\left(\frac{1}{n}\right) \approx 2.053755 \quad (27)$$

which means that the asymptotic variance of $L_2$ converges to a constant and is bounded away from zero.

### 8.5. *Mean of $T(g)$*

If $g(Z_{ij}) = g\left(|X_i - Y_i|\right)$ is a symmetric function, Taylor expansion leads to:

$$\begin{aligned} ET &= E\left(g\left(|X_i - Y_i|\right)\right) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(|x-y|\right) \\ &\times \left[\frac{1}{x^2+1} + \frac{2x}{(x^2+1)^2}\frac{h_x}{\sqrt{n}} + \frac{6x^2-2}{(x^2+1)^3}\frac{h_x^2}{2n} + O\left(\frac{1}{n}\right)\right] \\ &\times \left[\frac{1}{y^2+1} + \frac{2y}{(y^2+1)^2}\frac{h_y}{\sqrt{n}} + \frac{6y^2-2}{(y^2+1)^3}\frac{h_y^2}{2n} + O\left(\frac{1}{n}\right)\right]\frac{dx}{\pi}\frac{dy}{\pi} \end{aligned} \quad (28)$$

By existence of moments

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(|x-y|\right)\frac{1}{x^2+1}\frac{1}{y^2+1}\frac{dx}{\pi}\frac{dy}{\pi} = c_1 < \infty. \quad (29)$$

Further, let $r(x,y) = g\left(|x-y|\right)\frac{1}{x^2+1}\frac{2y}{(y^2+1)^2}$. Since $r(x,y) = r(-x,-y)$ and $r(x,y) = r(y,x)$,

$$\begin{aligned} \int^{+\infty}\int_0^{+\infty} r(x,y)dxdy &= \int_{-\infty}^{0}\int_{-\infty}^{0} r(x,y)dxdy, \\ \int_{-\infty}^{0}\int_0^{+\infty} r(x,y)dxdy &= \int_0^{+\infty}\int_{-\infty}^{0} r(x,y)dxdy, \end{aligned} \quad (30)$$

Then, we conclude

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(|x-y|\right)\frac{1}{x^2+1}\frac{2y}{(y^2+1)^2}dxdy = 0. \tag{31}$$

Then, $ET = c_1 + O\left(\frac{1}{n}\right)$. With calculation of the second derivative,

$$c_2 := \quad \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(|x-y|\right)\frac{1}{x^2+1}\frac{6y^2-2}{\left(y^2+1\right)^3}\frac{dx}{\pi}\frac{dy}{\pi} \tag{32}$$

a more accurate approximation becomes available

$$ET = c_1 + c_2\left(\frac{h_x^2}{2n}+\frac{h_y^2}{2n}\right) + O\left(n^{-3/2}\right).$$

### 8.6. *Variance of $T(g)$*

$C_x = Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_1j_2}\right)\right) = Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_1}\right)\right) = C_y,$

$$\lim_{n\to\infty} nVar\left(T\right) \quad = \quad (p_x^{-1}+p_y^{-1})C_x. \tag{33}$$

$$\begin{aligned}
C_x \;&=\; Cov\left[g\left(|x-z|\right), g\left(|y-z|\right)\right] = E\left[g\left(|x-z|\right)\cdot g\left(|y-z|\right)\right] - O(n^{-2}) \\
&=\; \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(|x-z|\right) g\left(|y-z|\right) \\
&\times\; \left[\frac{1}{x^2+1}+\frac{2x}{\left(x^2+1\right)^2}\frac{h_x}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\left[\frac{1}{y^2+1}+\frac{2y}{\left(y^2+1\right)^2}\frac{h_x}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right] \\
&\times\; \left[\frac{1}{z^2+1}+\frac{2z}{\left(z^2+1\right)^2}\frac{h_y}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\frac{dx}{\pi}\frac{dy}{\pi}\frac{dz}{\pi} + O(n^{-2}) \\
&=\; \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} g\left(|x-z|\right)\frac{1}{x^2+1}\frac{dx}{\pi}\right]^2\left[\frac{1}{z^2+1}+\frac{2z}{\left(z^2+1\right)^2}\frac{h_y}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\frac{dz}{\pi} \\
&+\; \frac{2h_y}{\sqrt{n}}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\left(|x-z|\right) g\left(|y-z|\right) \\
&\times\; \left[\frac{1}{x^2+1}\right]\left[\frac{2y}{\left(y^2+1\right)^2}\right]\frac{dx}{\pi}\frac{dy}{\pi}\frac{dz}{\pi} + O\left(\frac{1}{n}\right)
\end{aligned} \tag{34}$$

Note, the argument (30) applies and the second summand is equal to zero. Then, using

$$\int_{-\infty}^{+\infty} g(|x-z|)\frac{1}{x^2+1}\frac{dx}{\pi} = c_3(z) < \infty$$

for all $z$, and the equation simplifies to

$$C_x \quad = \quad \int_{-\infty}^{+\infty}\left[c_3(z)\right]^2\left[\frac{1}{z^2+1}+\frac{2z}{\left(z^2+1\right)^2}\frac{h_y}{\sqrt{n}}+O\left(\frac{1}{n}\right)\right]\frac{dz}{\pi}. \tag{35}$$

Numeric integration can be used to find

$$c_4 := \int_{-\infty}^{+\infty} [c_3(z)]^2 \left[ \frac{1}{z^2+1} \right] \frac{dz}{\pi}. \tag{36}$$

and, applying (30) arguments,

$$\int_{-\infty}^{+\infty} [c_3(z)]^2 \left[ \frac{2z}{(z^2+1)^2} \right] \frac{dz}{\pi} = 0 \tag{37}$$

leading to

$$C_x \quad = \quad c_4 + O\left(\frac{1}{n}\right) \tag{38}$$

which means that the asymptotic variance of $T$ defined by (33) converges to a constant and is bounded away from zero.

### *8.7. COVID-19: three weeks of the explosive phase of the epidemic*

[Table 4 about here.]

[Table 5 about here.]

## References


Aslan, B. and G. Zech (2005). New test for the multivariate two-sample problem based on the concept of minimum energy. *Journal of Statistical Computation and Simulation 75*(2), 109–119.

Chung, E. and J. P. Romano (2016). Asymptotically valid and exact permutation tests based on two-sample u-statistics. *Journal of Statistical Planning and Inference 168*, 97–105.

Chung, E., J. P. Romano, et al. (2013). Exact and asymptotically robust permutation tests. *The Annals of Statistics 41*(2), 484–507.

Hoeffding, W. (1948). A non-parametric test of independence. *The annals of mathematical statistics*, 546–557.

Keller-Mcnulty, A. and J. J. Higgins (1987). Effect of tail weight and outliers on power and type-i error of robust permutation tests for location: Effect of tail weight and outliers on power and type-i error. *Communications in Statistics-Simulation and Computation 16*(1), 17–35.

Le Cam, L. M. (1960). Locally asymptotically normal families of distributions. *Univ. California Publ. Statist. 3*, 37–98.

Lehmann, E. (1986). *Testing Statistical Hypotheses*. Probability and Statistics Series. Wiley.

Mann, H. B. and D. R. Whitney (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.

Marozzi, M. (2004). Some remarks about the number of permutations one should consider to perform a permutation test. *Statistica 64*(1), 193–201.

Nikitin, Y. Y. (1995). *Asymptotic efficiency of nonparametric tests*. Cambridge University Press.

Phillips, P. C., Y. Wu, and J. Yu (2009). Explosive behavior and the nasdaq bubble in the 1990s: when did irrational exuberance escalate asset values. *Int Econ Rev, in press. Also available at SSRN: http://ssrn. com/abstract 1413830*.

Pitman, E. J. (1949). Notes on non-parametric statistical inference. Technical report, North Carolina State University. Dept. of Statistics.

Rempala, G. A. and S. W. Looney (2006). Asymptotic properties of a two sample randomized test for partially dependent data. *Journal of statistical planning and inference 136*(1), 68–89.

Uchaikin, V. V. and V. M. Zolotarev (2011). *Chance and stability: stable distributions and their applications*. Walter de Gruyter.

Weinberg, J. M. and S. W. Lagakos (2000). Asymptotic behavior of linear permutation tests under general alternatives, with application to test selection and study design. *Journal of the American Statistical Association 95*(450), 596–607.

White, J. S. (1958). The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics*, 1188–1197.

Zhu, C., R. Byrd, P. Lu, and J. Nocedal (1995). A limited memory algorithm for bound constrained optimisation. *SIAM J. Sci. Stat. Comput 16*(5), 1190–1208.

**Table 1.** Power properties: Cauchy distribution with location = $4/\sqrt{n_x}$, scale = 1 and $n_x = n_y$. $L_{2,perm}$ for $n_x > 80$ were not calculated due to high computational cost

| $n_x$ | $h/\sqrt{n_x}$ | $L_2$ | $L_{2,perm}$ | $WMW$ | $T$ | $LRT$ | $GLRT$ |
|---|---|---|---|---|---|---|---|
| 20 | 0.89 | 7.2 | 30.8 | 28.4 | 5.4 | 85.5 | 43.2 |
| 50 | 0.57 | 5.7 | 33.3 | 31.5 | 4.2 | 87.5 | 48.4 |
| 80 | 0.45 | 5.2 | 35.6 | 34.4 | 2.6 | 88.0 | 51.1 |
| 160 | 0.32 | 6.0 | | 33.5 | 2.1 | 88.5 | 49.9 |
| 300 | 0.23 | 5.4 | | 34.1 | 3.2 | 85.6 | 52.3 |
| 500 | 0.18 | 3.4 | | 37.1 | 1.3 | 88.6 | 54.8 |

**Table 2.** Power (%) of permutation $L_\gamma$, LRT for Normal ($lrt_n$), Cauchy ($lrt_c$), Levy ($lrt_l$), and log-Cauchy ($lrt_{lc}$) distributions, Wilcoxon-Mann-Whitney (WMW) and Kolmogorov-Smirnov tests (KS); $n_x = x_y = 50$.

| $F_1$ | $F_2$ | $L_{0.5}$ | $L_1$ | $L_2$ | $lrt_n$ | $lrt_c$ | $lrt_l$ | $lrt_{lc}$ | WMW | KS |
|---|---|---|---|---|---|---|---|---|---|---|
| N(0, 1) | N(0, 1) | 4.9 | 5 | 4.3 | 5.3 | 5.1 | 5.2 | | 4.1 | 4.1 |
| N(0, 1) | N(0.6, 1) | 71.8 | 75.3 | 78.9 | 75 | 62.2 | 19.2 | | 81.7 | 68.4 |
| N(0, 1) | N(0, 2) | 88.6 | 89.3 | 90.3 | 99.1 | 69.8 | 75.5 | | 5.5 | 37.7 |
| N(0, 1) | N(0.5, 1.5) | 67.5 | 71 | 72.5 | 88.2 | 51.7 | 16.8 | | 45.9 | 50.7 |
| N(0, 1) | N(0.5, 2) | 94.3 | 95.1 | 95.4 | 99.6 | 81.4 | 56.1 | | 33.3 | 69.1 |
| C(0, 1) | C(0, 1) | 4.7 | 4.5 | 4.5 | 4.7 | 4.2 | 6.1 | | 5.5 | 4.1 |
| C(0, 1) | C(1, 1) | 81.3 | 80.8 | 79.1 | 5.9 | 85.3 | 5.2 | | 72.2 | 80.2 |
| C(0, 1) | C(0, 3) | 91.1 | 91.2 | 90.9 | 33.8 | 93.7 | 24.5 | | 6.4 | 45.9 |
| C(0, 1) | C(1, 2) | 82.2 | 81.9 | 80.7 | 17.2 | 85.3 | 11.8 | | 46.1 | 65.6 |
| C(0, 1) | C(1, 3) | 96.2 | 96.4 | 96.4 | 31 | 97.8 | 21.8 | | 28.8 | 70.4 |
| L(0, 1) | L(0, 1) | 5.1 | 5.3 | 5.3 | 5.6 | 5 | 4.3 | | 4.6 | 3.4 |
| L(0, 1) | L(1, 1) | 81.2 | 66.1 | 50.9 | 4.4 | 39.6 | 100 | | 52.9 | 92.3 |
| L(0, 1) | L(0, 3) | 74.6 | 72.3 | 70.8 | 11.3 | 69.3 | 92.7 | | 85.5 | 76.2 |
| L(0, 1) | L(0.5, 1.5) | 43.4 | 34.3 | 29.6 | 5.9 | 27.5 | 98.9 | | 50.5 | 58.6 |
| L(0, 1) | L(0.5, 2) | 65.5 | 56.4 | 51.6 | 6.1 | 48.8 | 99.5 | | 73.1 | 78 |
| LC(0, 1) | LC(0, 1) | 4.4 | 4.3 | 4.2 | | 6 | | 6.1 | 6.5 | 5.1 |
| LC(0, 1) | LC(1, 1) | 65.9 | 65 | 64.8 | | 87.7 | | 86.7 | 71.9 | 80.4 |
| LC(0, 1) | LC(0, 3) | 61.5 | 55.9 | 52.6 | | 41.4 | | 94.4 | 5.9 | 45.8 |
| LC(0, 1) | LC(1, 2) | 65.1 | 64.5 | 64 | | 72.3 | | 86 | 44.2 | 66.9 |
| LC(0, 1) | LC(1, 3) | 79.9 | 79.1 | 78.1 | | 67.9 | | 97.2 | 28.5 | 70.8 |

| Choice | Number of Elements | $Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_2}\right)\right)$ | i.i.d. case |
|---|---|---|---|
| $i_1 \neq i_2$ and $j_1 \neq j_2$ | $n_x^2n_y^2 - n_x^2n_y - n_xn_y^2 + n_xn_y$ | $0$ | $0$ |
| $i_1 = i_2$ and $j_1 \neq j_2$ | $n_y(n_x^2 - n_x)$ | $Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_1j_2}\right)\right)$ | $C_x$ |
| $i_1 \neq i_2$ and $j_1 = j_2$ | $n_x(n_y^2 - n_y)$ | $Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_1}\right)\right)$ | $C_y$ |
| $i_1 = i_2$ and $j_1 = j_2$ | $n_xn_y$ | $Var\left(g\left(Z_{i_1j_1}\right)\right)$ | $V$ |

**Table 3.** Summary of $Cov\left(g\left(Z_{i_1j_1}\right), g\left(Z_{i_2j_2}\right)\right)$ elements

| $\hat{\beta}_0$ | $\hat{\beta}_1$ | Country | Day0 | $\hat{\beta}_0$ | $\hat{\beta}_1$ | Country | Day0 |
|---|---|---|---|---|---|---|---|
| 1.2969 | -0.0094 | China | 1/22 | 1.2254 | -0.0024 | Brazil | 3/21 |
| 1.284 | -0.0048 | Iran | 2/25 | 1.2623 | -0.0105 | Denmark | 3/22 |
| 1.5033 | -0.0156 | Italy | 2/26 | 1.1866 | -0.0077 | Greece | 3/22 |
| 1.1811 | -0.0074 | South Korea | 2/26 | 1.2722 | -0.0106 | Portugal | 3/22 |
| 1.1785 | 0.008 | United States | 3/5 | 1.3641 | -0.0133 | Turkey | 3/22 |
| 1.4472 | -0.0121 | France | 3/8 | 1.2135 | -0.0074 | Austria | 3/23 |
| 1.5417 | -0.0196 | Spain | 3/10 | 1.2562 | -0.0091 | Ecuador | 3/23 |
| 1.1137 | -0.0041 | Japan | 3/11 | 1.1356 | -8.00E-04 | Egypt | 3/23 |
| 1.3113 | -0.0047 | United Kingdom | 3/14 | 1.1757 | -0.008 | Malaysia | 3/23 |
| 1.4083 | -0.0147 | Netherlands | 3/15 | 1.2029 | -0.004 | Hungary | 3/25 |
| 1.3848 | -0.0142 | Switzerland | 3/15 | 1.1923 | -0.0061 | Norway | 3/25 |
| 1.4359 | -0.0144 | Germany | 3/16 | 1.3393 | -0.0132 | Poland | 3/25 |
| 1.1606 | -0.0022 | Philippines | 3/16 | 1.2786 | -0.0109 | Romania | 3/25 |
| 1.1505 | -0.0058 | Iraq | 3/18 | 1.1986 | -0.0091 | Australia | 3/26 |
| 1.0805 | -0.0031 | San Marino | 3/18 | 1.1866 | -0.006 | Dominican Republic | 3/26 |
| 1.2167 | -0.0023 | Belgium | 3/19 | 1.3199 | -0.0105 | India | 3/26 |
| 1.197 | -9.00E-04 | Sweden | 3/19 | 1.1997 | -0.0077 | Argentina | 3/27 |
| 1.3392 | -0.0089 | Canada | 3/20 | 1.217 | -0.0068 | Ireland | 3/27 |
| 1.1898 | -0.007 | Indonesia | 3/20 | 1.2298 | -0.0108 | Morocco | 3/27 |
| 1.3599 | -0.0124 | Algeria | 3/21 | | | | |

**Table 4.** $\hat{\beta}_0$ and $\hat{\beta}_1$ of countries with EARLY outbreak of COVID-19 epidemic

| $\hat{\beta}_0$ | $\hat{\beta}_1$ | Country | Day0 | $\hat{\beta}_0$ | $\hat{\beta}_1$ | Country | Day0 |
|---|---|---|---|---|---|---|---|
| 1.2121 | -0.0076 | Israel | 3/28 | 1.1488 | -0.0065 | Macedonia | 4/2 |
| 1.1537 | -0.0068 | Luxembourg | 3/28 | 1.1833 | -0.0087 | Puerto Rico | 4/2 |
| 1.3249 | -0.0111 | Mexico | 3/28 | 1.1378 | -0.0072 | Tunisia | 4/2 |
| 1.1438 | -0.0045 | Panama | 3/28 | 1.113 | -0.0043 | Estonia | 4/3 |
| 1.2186 | -0.0069 | Peru | 3/28 | 1.0881 | -0.0019 | Bolivia | 4/4 |
| 1.0915 | -0.0045 | Albania | 3/29 | 1.0543 | -0.0033 | Cyprus | 4/4 |
| 1.2449 | -0.0112 | Czech Republic | 3/29 | 1.1339 | -0.0055 | Croatia | 4/5 |
| 1.1531 | -0.0035 | Pakistan | 3/29 | 1.1399 | -0.0043 | Moldova | 4/5 |
| 1.1599 | -0.0057 | Serbia | 3/29 | 1.0811 | 0.001 | United Arab Emirates | 4/5 |
| 1.2516 | -0.0095 | Colombia | 3/30 | 1.1098 | -0.0047 | Lithuania | 4/6 |
| 1.1279 | -0.0023 | Finland | 3/30 | 1.0275 | 0.0021 | Niger | 4/6 |
| 1.085 | -0.0049 | Lebanon | 3/30 | 1.1199 | -0.0021 | South Africa | 4/6 |
| 1.1606 | -0.0061 | Slovenia | 3/30 | 1.2493 | -0.0106 | Bangladesh | 4/7 |
| 1.0874 | -0.0029 | Burkina Faso | 3/31 | 1.1297 | -0.004 | Belarus | 4/7 |
| 1.1877 | -0.0016 | Russia | 3/31 | 1.0921 | -0.0013 | Afghanistan | 4/8 |
| 1.168 | -0.0086 | Thailand | 3/31 | 1.1181 | -0.0033 | Cuba | 4/8 |
| 1.2194 | -0.0078 | Ukraine | 3/31 | 1.1672 | -0.0069 | Cameroon | 4/9 |
| 1.1058 | -0.0048 | Andorra | 4/1 | 1.1084 | -0.0051 | Armenia | 4/10 |
| 1.1172 | -0.0053 | Bosnia & Herzegovina | 4/1 | 1.0715 | -0.0028 | Azerbaijan | 4/11 |
| 1.1939 | -0.007 | Chile | 4/1 | 1.0937 | -0.0046 | Kazakhstan | 4/11 |
| 1.1051 | -0.0034 | Honduras | 4/1 | 1.0374 | 0.0051 | Nigeria | 4/12 |
| 1.1568 | -0.0053 | Saudi Arabia | 4/1 | 1.08 | -0.003 | Mali | 4/14 |
| 1.1239 | -0.005 | Bulgaria | 4/2 | 1.0117 | 0.0016 | Singapore | 4/15 |
| 1.0728 | -0.0035 | Dem. Rep. of Congo | 4/2 | | | | |

**Table 5.** $\hat{\beta}_0$ and $\hat{\beta}_1$ of countries with LATE outbreak of COVID-19 epidemic

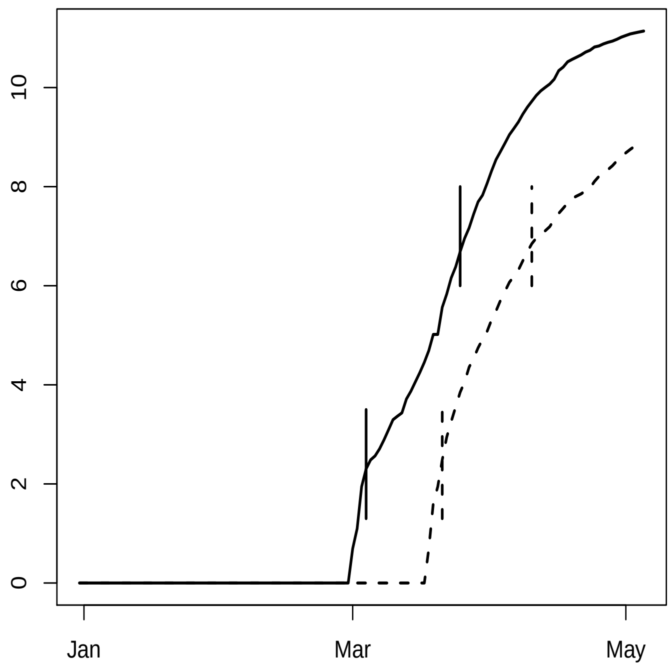



**Figure 1.** Daily numbers of total COVID-19 related deaths in the USA (solid line) and Brazil (dashed line) with marked 21 day periods starting with the first day of 10+ deaths.

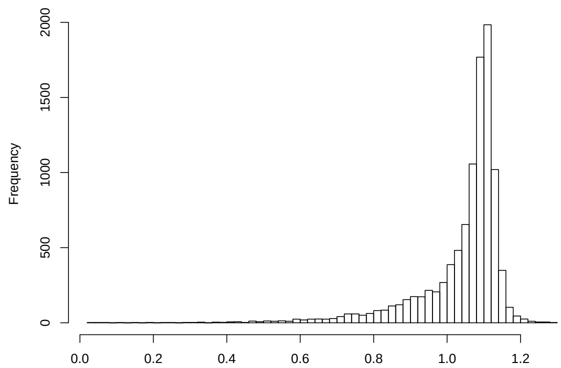
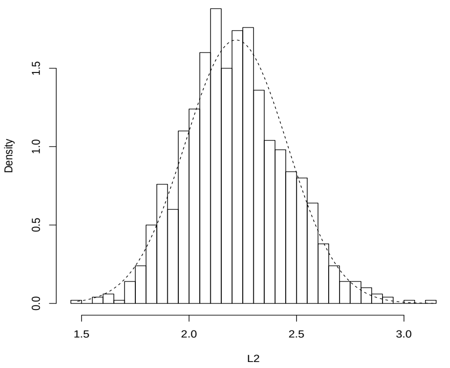
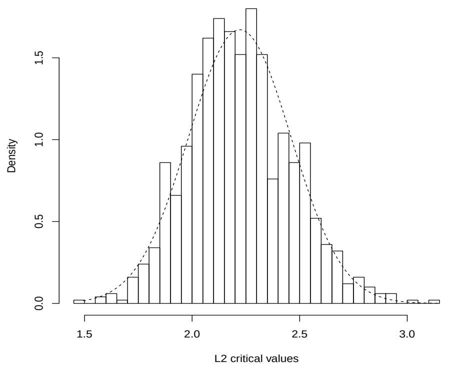


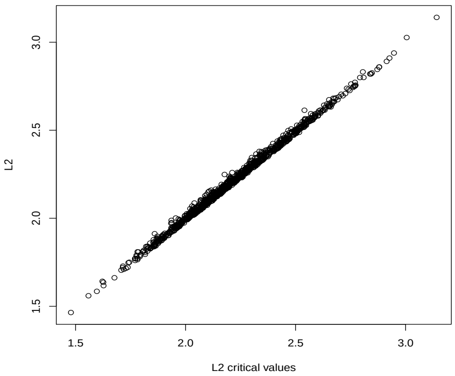

**Figure 2.** Simulated distribution of $\hat{\delta}$ with $\delta = 1.1$ in $Y_t = \delta Y_{t-1} + \epsilon_t, \epsilon_t \sim N(0,1), 1 < t \leq 21$.



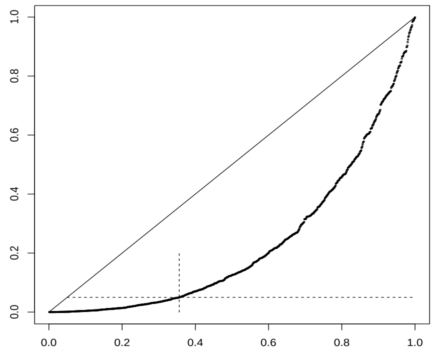

(a) Histogram of $L_2$



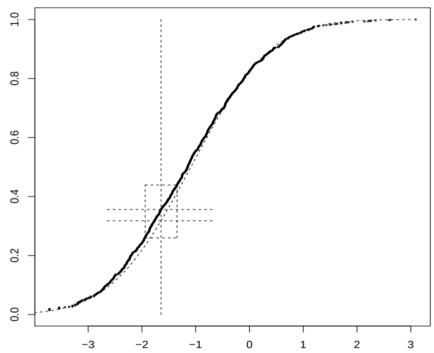

(b) Histogram of Critical values of $L_2$



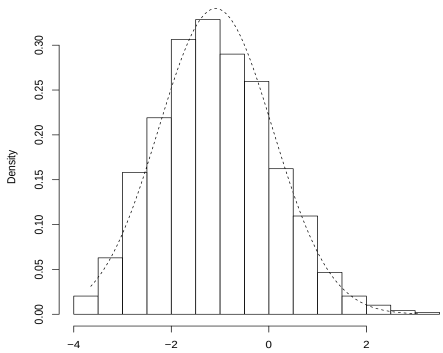

(c) Scatterplot of $L_2$ and its critical values



(d) Scatterplot of permutation P-values of $L_2$

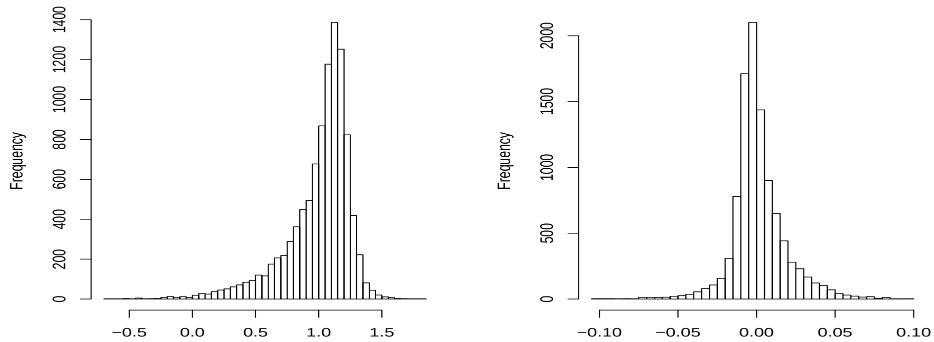


(e)




(f)

**Figure 3.** Monte-Carlo simulation results; $1,000$ resamples, $h = 4$, $n_x = n_y = 80$, $X \sim Cauchy(h/\sqrt{n}, 1)$




(a) Distribution of $\hat{\beta}_0$

(b) Distribution of $\hat{\beta}_1$

**Figure 4.** Simulated distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ in $Y_t = (\beta_0 + \beta_1 \cdot t)\, Y_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$, $\beta_0 = 1.1$, $\beta_1 = -0.001$, $1 < t \leq 21$.

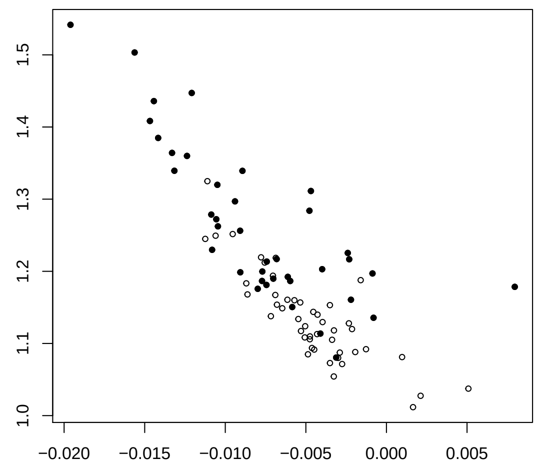



**Figure 5.** Scatterplot of $\hat{\beta}_{i0}$ and $\hat{\beta}_{i1}$; early entry into COVID-19 pandemic are drawn by solid circles, late entries are presented by the empty circles