

TEST FOR CHANGES IN THE MODELED SOLVENCY CAPITAL REQUIREMENT OF AN INTERNAL RISK MODEL

BY

DANIEL GAIGALL

ABSTRACT

In the context of the Solvency II directive, the operation of an internal risk model is a possible way for risk assessment and for the determination of the solvency capital requirement of an insurance company in the European Union. A Monte Carlo procedure is customary to generate a model output. To be compliant with the directive, validation of the internal risk model is conducted on the basis of the model output. For this purpose, we suggest a new test for checking whether there is a significant change in the modeled solvency capital requirement. Asymptotic properties of the test statistic are investigated and a bootstrap approximation is justified. A simulation study investigates the performance of the test in the finite sample case and confirms the theoretical results. The internal risk model and the application of the test is illustrated in a simplified example. The method has more general usage for inference of a broad class of law-invariant and coherent risk measures on the basis of a paired sample.

KEYWORDS

Bootstrap, Empirical process, Functional Delta method, Hadamard differentiability, Paired sample

1. INTRODUCTION

Solvency II directive intends the unification of the regulation of insurance companies in the European Union. Acts like the German VaG (2016) put the directive into national law. For the determination of the solvency capital requirement, §27 VaG (2016) intends the application of an internal risk model or a (so-called) standard formula in an insurance company. Major

market players, in particular large joint-stock companies, decided to use an internal risk model. Requests on the internal risk model are not limited to the modeling of the solvency capital requirement, but provide the modeling of a whole forecast distribution of the own funds of the insurance company, see §116 VaG (2016). According to §97 VaG (2016), the modeled solvency capital requirement based on the Value-at-Risk at level 99.5% of the modeled forecast distribution of the own funds, where the time horizon is one year. In this regard, suitable definitions of the solvency capital requirement are given in Christiansen and Niemeyer (2014).

The process of risk assessment with an internal risk model can be seen as the application of a company specific deterministic function, which models the current asset and liability portfolio of the corporate, to certain risk factors, which model the framework conditions for the insurance company with a forecast horizon of one year. Due to the complexity of the process of risk assessment with an internal risk model, the modeled forecast distribution is unknown. Consequently, the modeled solvency capital requirement is not available. A Monte Carlo approach is customary to solve this problem. Here, samples from the distributional model of the risk factors, calibrated on historical data or expert judgments, work as input of the model and yield an output sample, and finally an empirical forecast distribution of the own funds. A related empirical solvency capital requirement is derived.

The risk management has to deal with two sources of uncertainty. The first one is the model uncertainty, which amounts for the modeling of a single IBM stock already approximately 100%, see Danielsson et al. (2016). For a discussion on model uncertainties and their management, see Stahl (2016). The second one is the Monte Carlo noise. Although the Monte Carlo noise is probably small compared with the general model uncertainty, the number of simulations is crucial to get stable numerical results, especially in insurance context where the distributions exhibit usually fat tails and the high level of 99.5% in the quantile leads to noticeable statistical fluctuations in the empirical solvency capital requirement. In principle, it is possible to eliminate the Monte Carlo noise more and more by increasing the number of Monte Carlo replications. In practice, the number of Monte Carlo replications is often limited by the capacities of the company in terms of computing time or operation processes, see Casarano et al. (2017). Here, specific approximation techniques for the forecast of the own funds in the field of life insurance, such as the replicating portfolio approach, require very computing time-intensive valuation methods. In addition, the manageability of the output of the simulation is very important for the smooth course of the operation processes in the company.

To be compliant with §120 VaG (2016), the internal risk model has to be validated with the aim of verification and checking the adequacy. See Dacorogna (2017) for an overview over approaches and techniques to validate internal risk model results. An important validation tool is statistical variation analysis of the output of different model runs, at least on the occasion of the yearly required reporting obligations in accordance with §251 VaG (2016), but also

in the course of model changes for instance. Here, the effect of calibration updates, changes in the asset or liability portfolio, methodological changes, or economic shocks can be detected and compared with the expectations. Because the empirical solvency capital requirement depends substantially on the seed as the source of the randomness in the Monte Carlo approach, see Culver Heitmann, and Weiß (2018), it is useful to save the seed for all model runs and to use the same seed for different model runs for validation purposes. For this purpose, an appropriate implementation of the Monte Carlo simulation is requested to obtain corresponding results for the empirical own funds for different model runs.

In the course of the model validation, it is evident to check whether there is a significant change in the modeled solvency capital requirement, for example, for the categorization of the materiality of a model change or an adjustment of the solvency capital requirement. For this purpose, we suggest a test for checking whether there is a significant change in the modeled solvency capital requirement in Section 2. In the examples mentioned, the test's level controls the probability that the model change is falsely categorized as nonmaterial or the probability that the solvency capital requirement won't be adjusted when it is necessary. The requested implementation of the Monte Carlo approach with a fixed seed for all model runs serves as a paired sample of empirical own funds from two model runs as data. If the implementation of the Monte Carlo procedure is inappropriate in the sense that it does not serve a paired sample as data, additional deliberations are necessary to clarify whether our testing approach is suitable.

In mathematical terms, we treat the testing problem whether or not the absolute value of the relative deviation between two quantiles exceeds a given level on the basis of a paired sample as data. There is a lot of literature about inference for two quantiles, where usually the two-sample case is considered and two independent samples from the underlying distributions are used as data. Often, normal distributions appear in the limits of the statistics under consideration. The testing problem of the equality of quantiles is treated in Kosorok (1999), Malekzadeh and Jafari (2018), Li et al. (2012), and Cox and Jaber (1985). Inference for the distance of two quantiles is the topic in Guo and Krishnamoorthy (2005), Ozturk and Balakrishnan (2009), Baklizi (2018), Bristol (1990), Albers and Löhnberg (1984), Kang, Kim, and Lee (2007), Chakraborti and Desu (2008), and Malekzadeh and Kharrati-Kopaei (2020). The ratios of quantiles are considered in Malekzadeh and Mahmoudi (2020), Huang and Johnson (2006), Huang (2016), and Huang (2016). A general approach to the two-sample problem on the basis of the quantiles is given in Prihoda (1981). Our inference based on a paired sample as data. Because the paired-samples case can be seen as a generalization of the two-sample case if the sample sizes for both samples coincide, our approach can be regarded as an extension in this sense.

In Section 2, we investigate asymptotic properties of the suggested test statistic and obtain a normal distribution in the limit. Here, empirical process

theory in Dudley (1984), van der Vaart & Wellner (1996), and Ziegler (1997) is combined with the concept of Hadamard differentiability and the functional Delta method in van der Vaart (1998) to obtain the asymptotic results. The results motivate a bootstrap procedure for the approximation of the standard deviation of the test statistic and finally for the determination of critical values. Using that the theory is available in a general setting of triangular arrays of random variables and a Glivenko–Cantelli result in Gänßler & Ziegler (1994), we obtain that the bootstrap procedure is suitable. In Section 3, a simulation study investigates the performance of the test for a finite number of Monte Carlo replications. It is seen that the test performs quite well. In addition, the outcome of the simulation study confirms the theoretical findings. An example for a simplified portfolio illustrates the internal risk model and the application of the test. Finally, Section 4 summarizes the results of the paper and outlines an extension of the method for inference of a broad class of law-invariant and coherent risk measures on the basis of a paired sample. Here, concepts of quasi Hadamard differentiability and a related functional Delta method such as in Beutner and Zähle (2010), Krätschmer, Schied, and Zähle (2015), and Beutner and Zähle (2016) apply. In addition, related research questions of interest are discussed.

2. TEST

At first, we introduce the model and state the testing problem. Then, we give the details for the Monte Carlo procedure under a fixed seed and suggest as well as investigate a reasonable test statistic. Finally, we justify a bootstrap procedure to obtain critical values.

2.1. Model and testing problem

Let us consider two model runs named with the numbers $k = 1, 2$. The source of the internal risk model in run $k = 1, 2$ is an input $(X_1^{(k)}, \dots, X_d^{(k)})$, where $X_1^{(k)}, \dots, X_d^{(k)}$ are $d \in \mathbb{N}$ risk factors, given by real-valued random variables. The joint distributions of these risk factors model the framework conditions for the insurance company with a forecast horizon of one year and are assumed to be known. The modeled forecast of the own funds of the insurance company in model run $k = 1, 2$ is obtained by the application of a company specific deterministic and measurable function $r^{(k)} : \mathbb{R}^d \rightarrow \mathbb{R}$, which models the current asset and liability portfolio of the corporate. In principle this function is known, but it is not generally given in explicit form. Due to the complexity of the internal risk model, it is too cumbersome to work with this function in practice. In what follows, we deal with this function as unknown. The forecast of the own funds in model run $k = 1, 2$ is modeled by the real-valued random variable

$$Y^{(k)} = r^{(k)}(X_1^{(k)}, \dots, X_d^{(k)}).$$

Reminding that the Value-at-Risk at level $(1 - \gamma) \in (0, 1)$ of a real-valued random variable Y is defined as

$$\text{VaR}_{1-\gamma}(Y) = \inf\{x \in \mathbb{R}; P(Y \leq x) \geq 1 - \gamma\},$$

we suppose that the distributions $F^{(k)}$ of $Y^{(k)}$, $k = 1, 2$, are absolutely continuous, where the related densities $f^{(k)}$ satisfy $f^{(k)}(\text{VaR}_\gamma(Y^{(k)})) > 0$. We define the modeled solvency capital requirement at level γ in model run $k = 1, 2$ as the Value-at-Risk at level $1 - \gamma$ of the difference of the (known) current own funds of the corporate $y^{(k)} \in \mathbb{R}$ and the modeled forecast of the own funds $Y^{(k)}$ in 1 year, that is

$$\text{SCR}^{(k)} = \text{VaR}_{1-\gamma}(y^{(k)} - Y^{(k)}) = y^{(k)} - \text{VaR}_\gamma(Y^{(k)}),$$

see, for example, Christiansen and Niemeyer (2014). For $\gamma = 0.5\%$ we obtain the modeled solvency capital requirement of the insurance company according to the Solvency II directive. We suppose that the modeled solvency capital requirement at level γ in model run $k = 1, 2$ does not vanish, and that the modeled solvency capital requirements at level γ for model runs 1 and 2 do not coincide.

We are interested in the relative change of the modeled solvency capital requirement from model run 1 to model run 2, defined by

$$\Delta = \left| \frac{\text{SCR}^{(2)} - \text{SCR}^{(1)}}{\text{SCR}^{(1)}} \right|.$$

Our previous assumptions on the solvency capital requirements in model runs 1 and 2 ensure that $\Delta \in (0, \infty)$. We aim to check whether there is a significant change in the modeled solvency capital requirement in the sense that the relative change of the modeled solvency capital requirement from model run 1 to model run 2 exceeds a given level $\delta \in (0, \infty)$, that is the testing problem

$$H_0 : \Delta \geq \delta \text{ versus } H_1 : \Delta < \delta.$$

It is H_0 the null hypothesis, H_1 the alternative hypothesis.

2.2. Monte Carlo approach and test statistic

As it is explained in the introduction, the modeled forecast distribution in model run $k = 1, 2$ is unknown, and therefore the modeled solvency capital requirement $\text{SCR}^{(k)}$ is unknown as well, and the same applies to the relative change of the modeled solvency capital requirement Δ . This problem is solved by a Monte Carlo simulation with $n \in \mathbb{N}$ replications. As it is explained in the introduction, it is useful to apply the same seed for the implementation of the Monte Carlo approach for both model runs. In this regard, an appropriate implementation of the Monte Carlo procedure is requested to obtain corresponding results for the empirical own funds in each of the

n Monte Carlo replications. A possible implementation of the Monte Carlo simulation starts with a set of independent random variables $U_{i,j}, i = 1, \dots, d, j = 1, \dots, n$, each with the same uniform distribution on the unit interval $(0, 1)$, which is the same for both model runs. Denoting by $C^{(k)} : [0, 1]^d \rightarrow \mathbb{R}$ the copula related to the joint distribution of the random vector $(X_1^{(k)}, \dots, X_d^{(k)})$, the (inverse) Rosenblatt transformation serves a random vector $(V_{1,j}^{(k)}, \dots, V_{d,j}^{(k)})$ depending only on $(U_{1,j}, \dots, U_{d,j})$ and with the joint distribution $C^{(k)}$ for each $j = 1, \dots, n$ and $k = 1, 2$. Denoting by $(F_i^{(k)})^{-1}$ the quantile function of the distribution of $X_i^{(k)}$ and setting $X_{i,j}^{(k)} = (F_i^{(k)})^{-1}(V_{i,j}^{(k)})$ for $i = 1, \dots, d, j = 1, \dots, n$, and $k = 1, 2$, the random vectors $(X_{1,j}^{(k)}, \dots, X_{d,j}^{(k)})$, $j = 1, \dots, n$, obtained in this way are independent and each with the same distribution as $(X_1^{(k)}, \dots, X_d^{(k)})$. Moreover, $(X_{1,j}^{(1)}, \dots, X_{d,j}^{(1)}, X_{1,j}^{(2)}, \dots, X_{d,j}^{(2)})$, $j = 1, \dots, n$, are independent and identically distributed random vectors. Defining the empirical own funds in model run $k = 1, 2$ by

$$Y_j^{(k)} = r^{(k)}(X_{1,j}^{(k)}, \dots, X_{d,j}^{(k)}),$$

$j = 1, \dots, n$, we have that

$$(Y_1^{(1)}, Y_1^{(2)}), \dots, (Y_n^{(1)}, Y_n^{(2)})$$

is a paired sample of independent bivariate random vectors, each with the same bivariate distribution function F , which will be used as data for the treatment of the testing problem. In particular, it is not assumed that the bivariate distribution F is the product measure of the related marginal distributions $F^{(1)} \otimes F^{(2)}$. Because $r^{(k)}$ is unknown, $k = 1, 2$, the underlying bivariate distribution F is unknown as well. To circumvent technical problems, we suppose that F is uniformly continuous. We point out that the presented implementation of the Monte Carlo approach serves independent and identically distributed bivariate random vectors as data and ensures the validity of our theoretical results stated below. If the implementation of the Monte Carlo procedure is inappropriate in the sense that it does not serve independent and identically distributed bivariate random vectors as data, additional deliberations are necessary to clarify whether our testing approach is suitable.

We obtain the empirical solvency capital requirement at level γ of the own funds in model run $k = 1, 2$ by

$$SCR_n^{(k)} = y^{(k)} - Y_{[\gamma n];n}^{(k)}$$

where $Y_{[\gamma n];n}^{(k)}$ is the sample quantile at level γ of the $Y_1^{(k)}, \dots, Y_n^{(k)}$. For $\gamma = 0.5\%$ we obtain the empirical solvency capital requirement of the insurance company according to the Solvency II directive. The relative change of the empirical

solvency capital requirement from model run 1 to model run 2 is defined by

$$\Delta_n = \left| \frac{\text{SCR}_n^{(2)} - \text{SCR}_n^{(1)}}{\text{SCR}_n^{(1)}} \right|.$$

Note that, from our previous assumptions, Δ_n takes almost surely values in $(0, \infty)$. We suggest the test statistic

$$T_n = \sqrt{n}(\Delta_n - \delta)$$

for the implementation of the test. We state the following result for our test statistic.

Theorem 1. *The following statements are valid.*

(a) *In the interior of the null hypothesis $\Delta > \delta$, the divergence of the test statistic holds almost surely*

$$T_n \longrightarrow +\infty \text{ as } n \rightarrow \infty.$$

(b) *On the boundary of the null hypothesis $\Delta = \delta$, the convergence in distribution of the test statistic holds*

$$T_n \xrightarrow{d} N \text{ as } n \rightarrow \infty,$$

where N is a real-valued random variable with a centered normal distribution and variance

$$\begin{aligned} \sigma^2 = & \gamma(1 - \gamma) \frac{(\text{SCR}^{(1)})^2 f^{(1)}(\text{VaR}_\gamma(Y^{(1)}))^2 + (\text{SCR}^{(2)})^2 f^{(2)}(\text{VaR}_\gamma(Y^{(2)}))^2}{f^{(1)}(\text{VaR}_\gamma(Y^{(1)}))^2 f^{(2)}(\text{VaR}_\gamma(Y^{(2)}))^2 (\text{SCR}^{(1)})^4} \\ & + 2(F(\text{VaR}_\gamma(Y^{(1)}), \text{VaR}_\gamma(Y^{(2)})) - \gamma^2) \\ & \times \frac{\text{SCR}^{(1)} \text{SCR}^{(2)}}{f^{(1)}(\text{VaR}_\gamma(Y^{(1)})) f^{(2)}(\text{VaR}_\gamma(Y^{(2)})) (\text{SCR}^{(1)})^4}. \end{aligned}$$

(c) *Under the alternative hypothesis $\Delta < \delta$, the divergence of the test statistic holds almost surely*

$$T_n \longrightarrow -\infty \text{ as } n \rightarrow \infty.$$

2.3. Bootstrap and testing procedure

Following the testing concept of Neyman and Pearson, the test has to be chosen such that the probability to reject the null hypothesis H_0 that there is a significant change in the solvency capital requirement falsely does not exceed the statistical significance level. Motivated by Theorem 1, small values of the test statistic should be statistically significant in the sense that we reject the null hypothesis H_0 if and only if the test statistic T_n does not exceed a critical value

$c_{n,\alpha}$ which depends on the given statistical significance level $\alpha \in (0, 1)$. Because we deal with a one-sided testing problem, where small values of the test statistic are statistically significant, it is reasonable to choose the critical value $c_{n,\alpha}$ as a quantile of order α of the distribution of the test statistic T_n on the boundary of the null hypothesis $\Delta = \delta$.

Unfortunately, the distribution of the test statistic T_n depends on the unknown underlying bivariate distribution of the output of the first and second model run F , even in the case that we are on the boundary of the null hypothesis $\Delta = \delta$. The same applies to the standard deviation of the distribution of the test statistic T_n , likewise asymptotically for σ . For that reason, related quantiles are not available as critical values. Motivated by Theorem 1, we use

$$\hat{c}_{n,\alpha} = \hat{\sigma}_n \Phi^{-1}(\alpha)$$

as critical value, where Φ^{-1} denotes the quantile function of the standard normal distribution $\mathcal{N}[0, 1]$, and $\hat{\sigma}_n$ is a suitable approximation for σ , the standard deviation of the distribution of the test statistic T_n on the boundary of the null hypothesis $\Delta = \delta$.

We suggest a bootstrap technique to obtain such an approximation. Under the premise that $\Delta = \delta$, the test statistic can be rewritten as

$$T_n = \sqrt{n}(\Delta_n - \Delta).$$

Now, let $(\hat{Y}_{1,n}^{(1)}, \hat{Y}_{1,n}^{(2)}, \dots, (\hat{Y}_{n,n}^{(1)}, \hat{Y}_{n,n}^{(2)})$ be a bootstrap sample obtained by n -times sampling with replacement from the original observations $(Y_1^{(1)}, Y_1^{(2)}), \dots, (Y_n^{(1)}, Y_n^{(2)})$. The bootstrap solvency capital requirement at level γ of the own funds in model run $k = 1, 2$ is

$$\widehat{\text{SCR}}_n^{(k)} = y^{(k)} - \hat{Y}_{[\gamma n]:n}^{(k)},$$

where $\hat{Y}_{[\gamma n]:n}^{(k)}$ is the sample quantile at level γ of the $\hat{Y}_{1,n}^{(k)}, \dots, \hat{Y}_{n,n}^{(k)}$. The relative change of the bootstrap solvency capital requirement capital from model run 1 to model run 2 is defined by

$$\hat{\Delta}_n = \left| \frac{\widehat{\text{SCR}}_n^{(2)} - \widehat{\text{SCR}}_n^{(1)}}{\widehat{\text{SCR}}_n^{(1)}} \right|$$

Note that, from our previous assumptions, $\hat{\Delta}_n$ takes almost surely values in $(0, \infty)$. Regarding the latter expression of the test statistic, the bootstrap version of the test statistic is given by

$$\hat{T}_n = \sqrt{n}(\hat{\Delta}_n - \Delta_n).$$

We state the following result for the bootstrap version of the test statistic.

Theorem 2. *In all cases $\Delta > \delta$, $\Delta = \delta$, or $\Delta < \delta$ we have the conditionally convergence in distribution of the bootstrap version of the test statistic*

$$\hat{T}_n \xrightarrow{d} N \text{ as } n \rightarrow \infty$$

given the original observations in probability, where N is a random variable with a centered normal distribution and variance

$$\begin{aligned} \sigma^2 = & \gamma(1 - \gamma) \frac{(\text{SCR}^{(1)})^2 f^{(1)}(\text{VaR}_\gamma(Y^{(1)}))^2 + (\text{SCR}^{(2)})^2 f^{(2)}(\text{VaR}_\gamma(Y^{(2)}))^2}{f^{(1)}(\text{VaR}_\gamma(Y^{(1)}))^2 f^{(2)}(\text{VaR}_\gamma(Y^{(2)}))^2 (\text{SCR}^{(1)})^4} \\ & + 2(F(\text{VaR}_\gamma(Y^{(1)}), \text{VaR}_\gamma(Y^{(2)})) - \gamma^2) \\ & \times \frac{\text{SCR}^{(1)} \text{SCR}^{(2)}}{f^{(1)}(\text{VaR}_\gamma(Y^{(1)})) f^{(2)}(\text{VaR}_\gamma(Y^{(2)})) (\text{SCR}^{(1)})^4}. \end{aligned}$$

In particular, the distribution of N is the same distribution as in the limit in Theorem 1 (b) if $\Delta = \delta$ is valid.

The stated conditionally convergence in distribution given the original observations in probability is in the sense of the the following convergence in probability:

$$\sup_{h \in \text{BL}_1(\mathbb{R})} |E(h(\hat{T}_n) | (Y_1^{(1)}, Y_1^{(2)}), \dots, (Y_n^{(1)}, Y_n^{(2)})) - E(h(N))| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty,$$

where $\text{BL}_1(\mathbb{R})$ denotes the set of all Lipschitz continuous functions $h : \mathbb{R} \rightarrow [-1, +1]$ with $|h(x) - h(y)| \leq |x - y|$ for all $x, y \in \mathbb{R}$, compare with Chapter 23 in van der Vaart (1998). It is important to note that the conditional distribution of the bootstrap version of the test statistic \hat{T}_n given the original observations is known. Motivated by Theorem 2, we use the standard deviation of the conditional distribution of the bootstrap version of the test statistic \hat{T}_n given the original observations $(Y_1^{(1)}, Y_1^{(2)}), \dots, (Y_n^{(1)}, Y_n^{(2)})$, denoted by $\hat{\sigma}_n$, as approximation for σ , the standard deviation of the distribution of the test statistic T_n on the boundary of the null hypothesis $\Delta = \delta$. In practice, values for $\hat{\sigma}_n$ are obtained by Monte Carlo simulation. Specifically, we suggest to reject the null hypothesis H_0 that there is a significant change in the solvency capital requirement if and only if $T_n \leq \hat{c}_{n,\alpha}$.

3. SIMULATION AND EXAMPLE

At first, we present a simulation study and discuss the expectations as well as the numerical results. Finally, an example for a simplified portfolio illustrates the internal risk model and the application of the test.

3.1. Simulation

The simulations are implemented with the statistical software R and serve empirical rejection rates as well as empirical distributions of the p -values of the test. The scope of the simulation study includes the investigation of the performance of the test for a finite number of Monte Carlo replications beyond the theoretical limit results stated in Section 2 and in a setting which reflects the situation in insurance practice in an appropriate way. For this purpose, we consider different situations under the null hypothesis and under the alternative hypothesis, where the distance to the null hypothesis, measured in terms of the relative deviation between δ and Δ , is -20% , -10% , 0% , $+10\%$, and $+20\%$. Moreover, the confirmation of the theoretical findings in Theorem 1 and in Theorem 2 is also in the scope of the simulation study. Therefore, we consider an increasing number of Monte Carlo replications of the internal model from $n = 2000$ to $n = 200,000$ as well.

Let us describe the choice of the distributions in the simulation study. In insurance practice the underlying distributions exhibit probably fat tails. We consider fat tailed distributions to represent insurance risks in an appropriate way. Specifically, for the distributions $F^{(1)}$ and $F^{(2)}$ of the forecast of the own funds $Y_i^{(1)}$ and $Y_i^{(2)}$, $i = 1, \dots, n$, we consider two different combinations. In the first combination, we choose $F^{(1)}$ as the Fréchet distribution $\text{Fre}(1, 1.5)$ with scale parameter 1 and shape parameter 1.5 as well as $F^{(2)}$ as the Fréchet distribution $\text{Fre}(1.5, 1.5)$ with scale parameter 1.5 and shape parameter 1.5. In this combination, we choose the current own funds as $y^{(1)} = \Gamma(\frac{1}{3})$ and $y^{(2)} = \frac{3}{2}\Gamma(\frac{1}{3})$, that are the expectations of the distributions $F^{(1)}$ and $F^{(2)}$. In the second combination, we choose $F^{(1)}$ as the Pareto distribution $\text{Par}(1, 1.5)$ with scale parameter 1 and shape parameter 1.5 as well as $F^{(2)}$ as the Pareto distribution $\text{Par}(1.5, 1.5)$ with scale parameter 1.5 and shape parameter 1.5. Here, we choose the current own funds also as the expectations of the distributions $F^{(1)}$ and $F^{(2)}$, that are $y^{(1)} = 3$ and $y^{(2)} = 4.5$. The dependency structure of the bivariate distribution F of $(Y_i^{(1)}, Y_i^{(2)})$, $i = 1, \dots, n$, is modeled by a copula $C(u, v)$, $u, v \in [0, 1]$. For the implementation, we use the R package copula, see The Comprehensive R Archive Network (2018). Consequently, the bivariate distribution function F is given by $F(x, y) = C(F^{(1)}(x), F^{(2)}(y))$ for $x, y \in \mathbb{R}$. Precisely, we consider two different classes of copulas. For the first class of copulas, we use the bivariate Gaussian copula $C(u, v) = C_{\text{normal}}(u, v) = \Phi_{2,R}(\Phi^{-1}(u), \Phi^{-1}(v))$, $u, v \in [0, 1]$, where $\Phi_{2,R}$ denotes the distribution function of the bivariate normal distribution with mean 0 and covariance matrix $R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, the correlation ρ is either 0.25 or 0.75, and Φ denotes the distribution function of the univariate standard normal distribution. For the second class of copulas, we use the bivariate t copula $C(u, v) = C_t(u, v) = t_{2,R}(t^{-1}(u), t^{-1}(v))$, $u, v \in [0, 1]$, where $t_{2,R}$ denotes the distribution function of the bivariate t distribution with location parameter 0 and shape matrix R defined as above, and t is the distribution function of the univariate standard t distribution. For details about the bivariate t distribution, see Devroye (1990).

Now, we explain the choice of the model parameters in the simulation study. To obtain the solvency capital requirement according to the Solvency II directive we choose the level $\gamma = 0.5\%$. As it is explained above, the behavior of the test in different situations under the null hypothesis and under the alternative hypothesis is of interest. For the investigation of this aspect, we choose the level δ subject to the relative change of the modeled solvency capital requirement Δ which arises with the bivariate distribution F of the own funds in model runs 1 and 2 and the current own funds $y^{(1)}$ and $y^{(2)}$ under consideration. Specifically, we choose different levels $\delta \in \{80\% \cdot \Delta, 90\% \cdot \Delta, 100\% \cdot \Delta, 110\% \cdot \Delta, 120\% \cdot \Delta\}$. Moreover, for the study of the behavior of the test for an increasing number of Monte Carlo replications of the internal model, we consider $n=2000$, $n=20,000$, as well as $n=200,000$, and note that the usual order of magnitude of the number of Monte Carlo replications of an internal model is four over even larger.

Let us describe how the empirical results of the the simulation study are obtained. During the implementation of the simulations, several quantities are approximated by Monte Carlo simulation. At first, we choose the bivariate distribution F of the own funds in model runs 1 and 2, the current own funds $y^{(1)}$ and $y^{(2)}$, and the number of Monte Carlo replications of the internal model n as it is explained above. Then, we determine the value of the relative change of the modeled solvency capital requirement Δ on the basis of 200,000,000 Monte Carlo replications and define the level δ as it is explained above. Now, we simulate 2000 replications of the testing procedure, where in each of the 2000 simulation runs the bootstrap estimator for the standard deviation $\hat{\sigma}_n$ is also approximated by 2000 Monte Carlo replications according to the sampling procedure described in Section 2. The empirical rejection rates at the statistical significance level $\alpha = 10\%$ of the 2000 replications of the testing procedure are displayed in Table 1. In addition, Figure 1 presents plots of the empirical distribution functions of the p -values of the 2000 replications of the testing procedure.

Now, we formulate our expectations on the simulation results. At first, we point out that the test statistic is strictly decreasing in δ . For that reason, the rejection rates of the test are increasing in δ and the p -values of the test are increasing in δ with respect to the usual stochastic order. For the level $\delta = 100\% \cdot \Delta$, we are on the boundary of the null hypothesis. Here, Theorem 1 (b) combined with Theorem 2 applies. For that reason, we expect that the empirical rejection rates are close to the significance level of 10%. Moreover, it is expected that the empirical distribution functions of the p -values of the test are close to the identity function, that is the distribution function of the uniform distribution on the unit interval. For the levels $\delta = 80\% \cdot \Delta$ and $\delta = 90\% \cdot \Delta$, we are in the interior of the null hypothesis. Now, Theorem 1 (a) combined with Theorem 2 applies. Consequently, we expect that the empirical rejection rates are less than the significance level of 10% and tend to zero as the number of Monte Carlo replications n increases. Furthermore, it is expected that the empirical distribution functions of the p -values of the test proceed

TABLE 1
EMPIRICAL REJECTION RATES FOR THE STATISTICAL SIGNIFICANCE LEVEL $\alpha = 10\%$.

F	n	δ				
		$80\% \cdot \Delta$	$90\% \cdot \Delta$	$100\% \cdot \Delta$	$110\% \cdot \Delta$	$120\% \cdot \Delta$
$C = C_{\text{normal}}, \rho = 0.25$	2000	0.0870	0.1140	0.1405	0.1675	0.1980
$F^{(1)} = \text{Fre}(1, 1.5), y^{(1)} = \Gamma(\frac{1}{3})$	20,000	0.0170	0.0515	0.1150	0.2015	0.3110
$F^{(2)} = \text{Fre}(1.5, 1.5), y^{(2)} = \frac{3}{2}\Gamma(\frac{1}{3})$	200,000	0.0000	0.0045	0.1140	0.5065	0.8800
$C = C_{\text{normal}}, \rho = 0.75$	2000	0.0705	0.0935	0.1205	0.1450	0.1795
$F^{(1)} = \text{Par}(1, 1.5), y^{(1)} = 3$	20,000	0.0155	0.0430	0.1165	0.2135	0.3495
$F^{(2)} = \text{Par}(1.5, 1.5), y^{(2)} = 4.5$	200,000	0.0000	0.0015	0.0915	0.5355	0.9275
$C = C_t, \rho = 0.75$	2000	0.0585	0.0835	0.1100	0.1425	0.1795
$F^{(1)} = \text{Fre}(1, 1.5), y^{(1)} = \Gamma(\frac{1}{3})$	20,000	0.0095	0.0335	0.0975	0.2235	0.4045
$F^{(2)} = \text{Fre}(1.5, 1.5), y^{(2)} = \frac{3}{2}\Gamma(\frac{1}{3})$	200,000	0.0000	0.0030	0.1155	0.6570	0.9755
$C = C_t, \rho = 0.25$	2000	0.0810	0.1100	0.1380	0.1695	0.1950
$F^{(1)} = \text{Par}(1, 1.5), y^{(1)} = 3$	20,000	0.0150	0.0370	0.1020	0.1985	0.3330
$F^{(2)} = \text{Par}(1.5, 1.5), y^{(2)} = 4.5$	200,000	0.0000	0.0035	0.0810	0.5025	0.9000

below the identity function and fall as the number of Monte Carlo replications of the internal model n increases. For the levels $\delta = 110\% \cdot \Delta$ and $\delta = 120\% \cdot \Delta$, we are under the alternative hypothesis. In this situation, Theorem 1 (c) combined with Theorem 2 applies. In this regard, we expect that the empirical rejection rates are greater than the significance level of 10% and tend to one as the number of Monte Carlo replications of the internal model n increases. In addition, it is expected that the empirical distribution functions of the p -values of the test proceed above the identity function and rise as the number of Monte Carlo replications n increases. Beyond that general expectations, the question how good the test really performs has to be investigated empirically.

Finally, we discuss the results of the simulation study. We point out that the outcome of the simulation study confirms all our expectations formulated above. The results for an increasing number of Monte Carlo replications of the internal model are in line with Theorem 1 and Theorem 2. Moreover, we find that the test performs quite well in a setting which reflects the situation in insurance practice in an appropriate way. At first, we consider the empirical rejection rates displayed in Table 1. Under the null hypothesis, the results obtained indicate that the test keeps the significance level quite well. Precisely, on the boundary of the null hypothesis $\delta = 100\% \cdot \Delta$ the empirical rejection rates are close to the statistical significance level of 10%, where the absolute deviations lie in a range between 0.2% and 4.05%. In the interior of the null hypothesis for levels $\delta \leq 90\% \cdot \Delta$ the empirical rejection rates are less than the statistical significance level of 10%, except in two cases for the lowest number of Monte Carlo replications $n = 2000$ where the excess is 1% and 1.4%, and tend strictly decreasing to zero as the number of Monte Carlo replications increases from $n = 2000$ to $n = 200,000$. It is also seen that the test has power under alternatives for levels $\delta \geq 110\% \cdot \Delta$. Precisely, the empirical rejection rates are

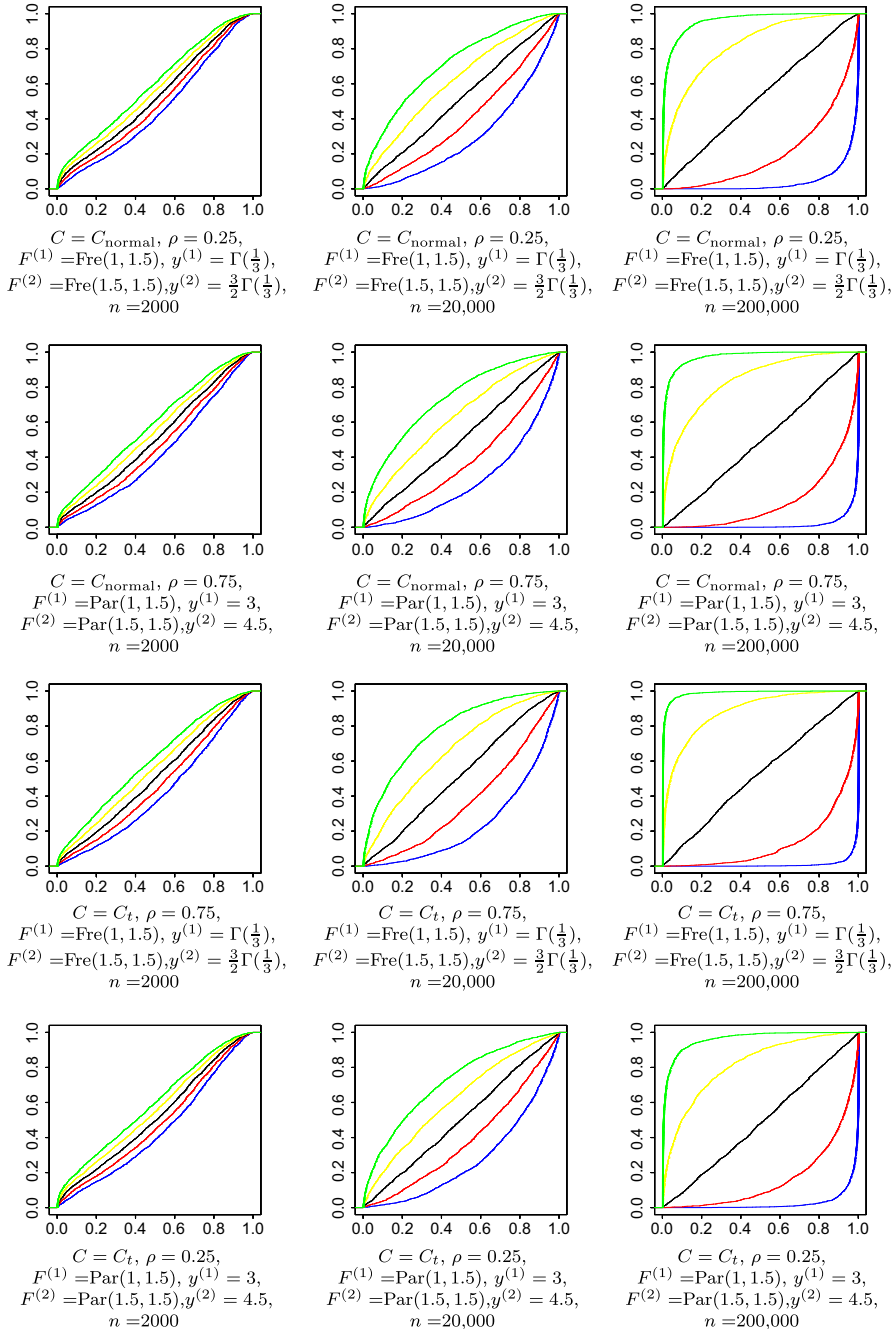


FIGURE 1: Empirical distribution functions of p-values for levels $\delta = 80\% \cdot \Delta$ (blue), $\delta = 90\% \cdot \Delta$ (red), $\delta = 100\% \cdot \Delta$ (black), $\delta = 110\% \cdot \Delta$ (yellow), and $\delta = 120\% \cdot \Delta$ (green).

clearly greater than the statistical significance level of 10%, or even close to one for $n=200,000$ and $\delta = 120\% \cdot \Delta$, and increase strictly as the number of Monte Carlo replications increases from $n=2000$ to $n=200,000$. Now, we consider the empirical distribution functions of the p -values of the test presented in Figure 1. The observation that the empirical distribution functions of the p -values of the test are close to the identity function for the level $\delta = 100\% \cdot \Delta$ shows that the bootstrap procedure serves a suitable approximation of the variance of the test statistic. In addition, the proceed of the empirical distribution functions of the p -values of the test for the other levels $\delta \leq 90\% \cdot \Delta$ or $\delta \geq 110\% \cdot \Delta$ shows that the test clearly distinguishes between situations where the null hypothesis or the alternative hypothesis is valid if the distance to the null hypothesis, measured in terms of the relative deviation between δ and Δ , is less that -10% or greater that $+10\%$.

3.2. Example

We consider a simplified portfolio consisting of an asset side and a liability side. The asset side is given by a single European put option on a stock, where the related market value is determined by the Black–Scholes pricing formula. The input of the asset side are economic scenarios of the risk categories equity and interest rate. The risk factor equity describes the evolution of the stock price and is modeled by a geometric Brownian motion. The risk factor interest rate follows the Cox–Ingersoll–Ross model. The risk factor of the liability side is given by the total amount of claims corresponding to a portfolio of simple policies. Here, we use the collective risk model, precisely a compound Poisson model, with a fat tailed distribution for the individual claims.

Specifically, for the evolution of the risk factors equity and interest rate, we consider the stochastic differential equations

$$dS_t = \mu S_t dt + \sigma_S S_t dW_t$$

and

$$dr_t = a(b - r_t)dt + \sigma_r \sqrt{r_t} d\tilde{W}_t,$$

which corresponds to a geometric Brownian motion and the Cox–Ingersoll–Ross model, respectively. Here, W_t and \tilde{W}_t denote independent Brownian motions. We set $\mu = 1$ for the stock drift, $\sigma_S = 1$ for the stock volatility, $a = 0.01$ for the speed of mean reversion parameter, $b = 0.02$ for the long term mean level, $\sigma_r = 1$ for the interest rate volatility, $S_0 = 1$ as start value for the stock, and $r_0 = 0.01$ as start value for the interest rate. The risk factor for the total amount of claims is given by

$$X_t = \sum_{i=1}^{N_t} Z_i,$$

where N_t is a Poisson process with intensity $\lambda = 1$, Z_1, Z_2, \dots are independent random variables each with the same Fréchet distribution $\text{Fre}(0.1, 1.5)$ with scale parameter 0.1 and shape parameter 1.5, and $W_t, \tilde{W}_t, N_t, Z_1, Z_2, \dots$ are independent. The portfolio, consisting of a single European put option and the total amount of claims, has the value $C(S_t, r_t, t) - X_t$, where

$$C(S, r, t) = \Phi(-d_2)K \exp(-r(T-t)) - \Phi(-d_1)S$$

is the value of the option, with

$$d_1 = \frac{1}{\sigma\sqrt{T-t}} \left(\log \frac{S}{K} + \left(r + \frac{\sigma_C}{2} \right) (T-t) \right), \quad d_2 = d_1 - \sigma\sqrt{T-t},$$

where $T = 2$ is the maturity and $K = 1.5$ is the strike price. Here, we denote by Φ the distribution function of the standard normal distribution $\mathcal{N}[0, 1]$. The current own funds and the related forecast distribution is obtained by evaluation of the portfolio value in $t = 0$ and $t = 1$. We consider the underlying volatility of the contract $\sigma_C = 1$. The example is also implemented with the statistical software R. Thereby, we fix the seed to 1 for both model runs. For the number of Monte Carlo replications of the internal model we choose $n = 20,000$. During the implementation of the example, several quantities are approximated by Monte Carlo simulation. We determine the value of the (relative change of the) modeled solvency capital requirement Δ on the basis of 2,000,000 Monte Carlo replications. Moreover, the bootstrap estimator for the standard deviation $\hat{\sigma}_n$ is approximated by 200,000 replications according to the sampling procedure described in Section 2. We choose the level $\gamma = 0.5\%$ for the solvency capital requirement according to the Solvency II directive. We treat parameter updates on the interest rate volatility $\sigma_r \in \{5, 10, 15, 20\}$ and on the volatility of the contract $\sigma_C \in \{5, 10, 15, 20\}$ in the second model run.

In what follows we discuss the results. Modeled and empirical solvency capital requirements and related relative changes are given in Table 2. We find noticeable deviations between the modeled and the related empirical quantities. The relative changes in the modeled and empirical solvency capital requirement lie in the range between 10% and 70%. Realized p -values of the test for levels $\delta = x \cdot \Delta$, $x \in (0\%, 200\%)$, are presented in Figure 2. Because the test statistic is strictly decreasing in δ , the realizations of the p -values are strictly decreasing in x . It is seen that the test starts to indicate a significant change in the modeled solvency capital requirement in terms of realized p -values less than 10% for δ 's which lie in the range between $\delta = 100\% \cdot \Delta$ and $\delta = 120\% \cdot \Delta$ for the volatilities σ_r and σ_C under consideration. Noticing that the power values of the test, obtained by simulation for $\alpha = 10\%$ and $n = 20,000$ in Subsection 3.1, lie in the range between 19.85% and 22.35% for the level $\delta = 110\% \cdot \Delta$ and in the range between 31.10% and 40.45% for the level $\delta = 120\% \cdot \Delta$, this result was not necessarily expected, but is not surprising.

TABLE 2
 MODELED AND EMPIRICAL SOLVENCY CAPITAL REQUIREMENTS AND RELATED RELATIVE CHANGES.

$\sigma_r = \sigma_C$	SCR	SCR ⁽¹⁾	SCR ⁽²⁾	Δ	SCR _n ⁽¹⁾	SCR _n ⁽²⁾	Δ_n
5		4.022875	4.647840	0.1553528	4.03318	4.583863	0.1365382
10		4.022875	5.977039	0.4857630	4.03318	5.855014	0.4517116
15		4.022875	6.526211	0.6222754	4.03318	6.317268	0.5663244
20		4.022875	6.803241	0.6911391	4.03318	6.512470	0.6147235

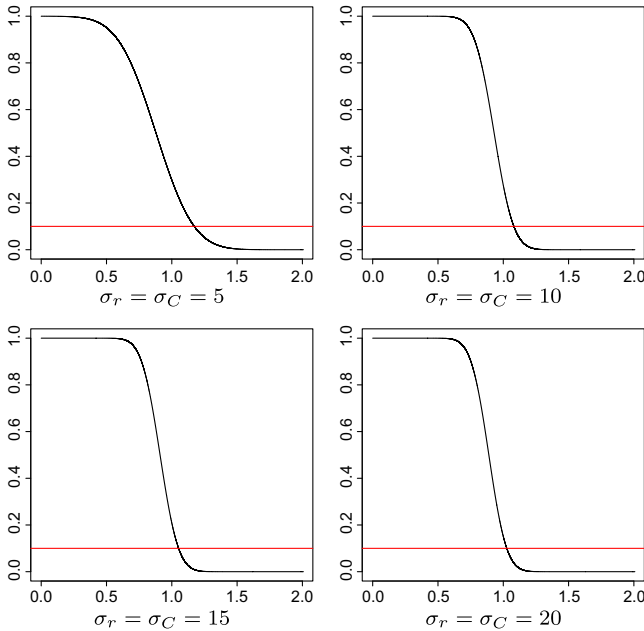


FIGURE 2: Plots x versus p -value of the test for level $\delta = x \cdot \Delta$, $x \in (0\%, 200\%)$, (black), and the statistical significance level $\alpha = 10\%$ (red).

4. CONCLUSION AND OUTLOOK

Finally, we summarize the main findings of the work in a conclusion, outline a possible extension of the idea, and discuss related research questions of interest.

4.1. Conclusion

The paper considers two model runs of an internal risk model and discusses the question of a significant change in the modeled solvency capital requirement. The problem is treated on the basis of a paired sample from the related distributions, served as the output of a Monte Carlo procedure under a fixed

seed for both model runs. A reasonable test statistic is suggested and the limit behavior of the test in different situations under the null hypothesis and under the alternative hypothesis is obtained. Due to the demand for critical values, a bootstrap procedure is suggested and justified. A simulation study investigates the performance of the test for a finite number of Monte Carlo replications beyond the theoretical results. The outcome of the simulation study confirms the expectations and the theoretical findings. In addition, we find that the test performs quite well in a setting which reflects the situation in insurance practice in an appropriate way. An example for a simplified portfolio consisting of an asset side and a liability side illustrates the internal risk model and how the test is applied.

4.2. Outlook

The method has more general usage for inference of general risk measures beyond the Value-at-Risk on the basis of a paired sample if the underlying risk measure is suitable. Here we define Δ and Δ_n with a more general risk measure ρ in the place of the Value-at-Risk. If ρ is Hadamard differentiable, our testing approach and our asymptotic results for the test statistic and the bootstrap procedure still apply in this situation. In the context of the Solvency II, the relevant risk measure is the Value-at-Risk and empirical process theory (for triangular arrays) is combined with the Hadamard differentiability of the quantile functional and the functional Delta method (for bootstrap) to obtain the asymptotic results. In fact, the request of Hadamard differentiability on the risk measure is a fairly strong condition. Nevertheless, Hadamard differentiability can be weakened to the so-called quasi Hadamard differentiability, see Krättschmer, Schied, and Zähle (2015), such that the concept applies to a broad class of law-invariant and coherent risk measures. In addition, the functional Delta method can be extended to functionals which satisfy this weaker assumption, see Beutner and Zähle (2010) and Beutner and Zähle (2016) for details. In all, it should be possible to adopt our argumentation to ensure the asymptotic properties of the test statistic on the basis of a broad class of law-invariant and coherent risk measures. Furthermore, the works Beutner and Zähle (2010) and Beutner and Zähle (2016) serve an extension of the functional Delta method for quasi Hadamard differentiable functionals such that the adoption of our argumentation for the bootstrap procedure should be possible for a broad class of law-invariant and coherent risk measures.

A related topic of interest, in particular on the side of the regulatory authority, is the detection of “model drift”, see Bank of England (2016). Regulators are concerned whether a change in the solvency capital requirement is driven by changes in the exposure or changes in the statistical assumptions. Thinking of an update in the drift parameter of the interest rate forecast, for instance, the parameters of the joint distribution of the risk factors $(X_1^{(k)}, \dots, X_d^{(k)})$ are determined on the basis of the statistical assumptions. Therefore, it is of interest

whether the reason for the change in the solvency capital requirement $SCR_n^{(k)}$ from model run $k = 1$ to $k = 2$ observed is a change in the joint distribution of the risk factors $(X_1^{(k)}, \dots, X_d^{(k)})$ given some information about the functions $r^{(k)}$. This is an important and interesting topic for a further research work.

ACKNOWLEDGMENTS

The author wishes to thank an Editor of *ASTIN Bulletin* and three referees for very detailed and helpful comments and suggestions. Special thanks goes to Gerhard Stahl, Executive Board of the HDI Service AG, Chief Risk Officer of the HDI Talanx Group and the HDI V.a.G, as well as Professor at Ulm University, for innumerable intellectual inspirations.

COMPETING INTERESTS DECLARATION

The author states that there are no competing interests.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/asb.2021.20>.

REFERENCES

- Albers, W. Löhnberg, P. (1984). An approximate confidence interval for the difference between quantiles in a bio-medical problem. *Statistica Neerlandica* 38, 20–22.
- Bank of England (2016). Monitoring model drift and standard formula SCR reporting for firms with an approved internal model. Supervisory Statement 15/16. <https://www.bankofengland.co.uk/prudential-regulation/publication/2016/solvency2-monitoring-model-drift-and-standard-formula-scr-reporting-ss>
- Baklizi, A. (2018). Interval Estimation of Quantile Difference in the Two-Parameter Exponential Distributions. *Journal of Testing and Evaluation* 46, 2654–2660.
- Beutner, E. Zähle, H (2010). A modified functional delta method and its application to the estimation of risk functionals. *Journal of Multivariate Analysis* 101, 2452–2463 .
- Beutner, E. Zähle, H (2016). Functional delta-method for the bootstrap of quasi Hadamard differentiable functionals. *Electronic Journal of Statistics* 10, 1181–1222.
- Bristol, D.R. (1990). Distribution-free confidence intervals for the difference between quantiles. *Statistica Neerlandica* 44, 87–90.
- Casarano, G., Castellani, G., Passalacqua, L., Perla, F., Zanetti, P. (2017). Relevant applications of Monte Carlo simulation in Solvency II. *Soft Computing* 21, 1181–1192.
- Chakraborti, S., Desu, M.M.(2008). A Distribution-Free Confidence Interval For The Difference Between Quantiles With Censored Data. *Statistica Neerlandica* 40, 93–98.
- Christiansen, M. C., Niemyer, A. (2014). The fundamental definition of the Solvency Capital Requirement in Solvency II. *ASTIN Bulletin* 44, 501–533.

- Culver, Q., Heitmann, D., Weiß, C. (2018). The Influence of Seed Selection on the Solvency II Ratio. *Der Aktuar* 1/2018.
- Cox, T.F., Jaber, K. (1985). Testing the Equality of two Normal Percentiles. *Communication in Statistics- Simulation and Computation* 14, 345–356.
- Casarano, G., Castellani, G., Passalacqua, L., Perla, F., Zanetti, P. (2017). Relevant applications of Monte Carlo simulation in Solvency II. *Soft Computing* 21, 1181–1192.
- Dacorogna, M. (2017). Approaches and Techniques to Validate Internal Model Results. *SSRN Electronic Journal*.
- Danielsson, J., James, K.R., Valenzuela, M., Zer, I. (2016). Model Risk of Risk Models. *Journal of Financial Stability* 23, 79–91.
- Devroye, L. (1990). A note on Linnik's distribution. *Statistics and Probability letters* 9, 305–306.
- Dudley, R. M. (1984). *A course on empirical processes*. Lecture Notes in Mathematics 1097, 1–142. Springer, New York.
- Gänßler, P., Ziegler, K. (1994). A Uniform Law of Large Numbers for Set-Indexed Processes with Applications to Empirical and Partial-Sum Processes. *Probability in Banach Spaces* 9, 385–400.
- Guo, H., Krishnamoorthy, K. (2005). Comparison Between Two Quantiles: The Normal and Exponential Cases. *Communication in Statistics- Simulation and Computation* 34, 243–252.
- Huang, L.-F., Johnson, R.A. (2006). Confidence regions for the ratio of percentiles. *Statistics & Probability Letters* 76, 384–392.
- Huang, L.-F. (2016). Approximated non parametric confidence regions for the ratio of two percentiles. *Communication in Statistics- Theory and Methods* 46, 4004–4015.
- Huang, L.-F. (2016). Adjusted Wilcoxon signed rank test tables for ratio of percentiles. *Communication in Statistics- Simulation and Computation* 46, 5763–5771.
- Li, X., Tian, L., Wang, J., Muindi, J.R. (2012). Comparison of quantiles for several normal populations. *Computational Statistics & Data Analysis* 56, 2129–2138.
- Kang, S.-G., Kim, D.-H., and Lee, W.-D. (2007). Noninformative Priors for the Difference of Two Quantiles in Exponential Models. *Communications for Statistical Applications and Methods* 14, 431–442.
- Kosorok, M.R. (1999). Two-Sample Quantile Tests under General Conditions. *Biometrika* 86, 909–921.
- Krätschmer, V., Schied, A., Zähle, H. (2015). Quasi-Hadamard differentiability of general risk functionals and its application. *Statistics & Risk Modeling* 32, 25–47.
- Malekzadeh, A., Jafari, A.A. (2018). Testing equality of quantiles of two-parameter exponential distributions under progressive Type II censoring. *Journal of Statistical Theory and Practice* 12, 776–793.
- Malekzadeh, A., Kharrati-Kopaei, M. (2020). Simultaneous confidence intervals for the quantile differences of several two-parameter exponential distributions under the progressive type II censoring scheme. *Journal of Statistical Computation and Simulation* 90, 1–20.
- Malekzadeh, A., Mahmoudi, S.M. (2020). Constructing a confidence interval for the ratio of normal distribution quantiles. *Monte Carlo Methods and Applications* 26, 325–334.
- Ozturk, O., Balakrishnan, N. (2009). Exact two-sample nonparametric test for quantile difference between two populations based on ranked set samples. *Annals of the Institute of Statistical Mathematics* 61, 235–249.
- Prihoda, T.J. (1981). A generalized approach to the two sample problem: the quantile approach. *Dissertation, Texas A&M University*.
- Stahl, G. (2016). Model Uncertainty in a Holistic Perspective. *Kallsen J., Papapantoleon A. (eds) Advanced Modelling in Mathematical Finance. Springer Proceedings in Mathematics & Statistics, vol 189. Springer, Cham, 189–215.*
- The Comprehensive R Archive Network. (2018). <https://cran.r-project.org/web/packages/copula/copula.pdf>
- VaG (2016). Gesetz über die Beaufsichtigung der Versicherungsunternehmen.
- van der Vaart, A., Wellner, J.A. (1996). *Weak convergence and empirical processes*. Springer, New York.
- van der Vaart. (1998). *Asyptotic Statistics*. Cambridge.

Ziegler, K. (1997). Functional Central Limit Theorems for Triangular Arrays of Function-Indexed Processes under Uniformly Integrable Entropy Conditions. *Journal of Multivariate Analysis* 62, 233–272.

DANIEL GAIGALL

Institute of Probability and Statistics,

Leibniz University Hannover,

Welfengarten 1, 30167 Hannover, Germany,

E-Mail: gaigall@stochastik.uni-hannover.de

House of Insurance, Leibniz University Hannover,

Welfengarten 1, 30167 Hannover, Germany,

E-Mail: daniel.gaigall@insurance.uni-hannover.de

Group Risk Management, HDI Service AG,

HDI-Platz 1, 30659 Hannover, Germany,

E-Mail: daniel.gaigall@hdi.de

APPENDIX A. PROOFS

At first, we prove Theorem 1. For the proof we use empirical process theory in Dudley (1984), van der Vaart & Wellner (1996), and Ziegler (1997) combined with the concept of Hadamard differentiability and the functional Delta method in van der Vaart (1998).

Proof. Proof of Theorem 1. At first, we prove (a) and (c) with standard arguments. From our assumptions on the distributions of the outcome of the model runs 1 and 2, we obtain the (joint) strong consistency for the sample quantiles the almost surely convergence

$$(Y_{[\gamma n]:n}^{(1)}, Y_{[\gamma n]:n}^{(2)}) \longrightarrow (q_{\gamma}^{(1)}, q_{\gamma}^{(2)}) \text{ as } n \rightarrow \infty,$$

where $q_{\gamma}^{(k)} = \text{VaR}_{\gamma}(Y^{(k)})$, $k = 1, 2$. Setting $\mathcal{Y} = \{(x, y) \in \mathbb{R}^2; y^{(1)} - x \neq 0, y^{(2)} - y \neq 0, y^{(1)} - x \neq y^{(2)} - y\}$, we define the map $\phi: \mathbb{R}^2 \rightarrow (0, \infty)$ by

$$\phi(x, y) = \left| \frac{y^{(2)} - y - (y^{(1)} - x)}{y^{(1)} - x} \right|, (x, y) \in \mathcal{Y},$$

and $\phi(x, y) = 0$, $(x, y) \in \mathbb{R}^2 \setminus \mathcal{Y}$. It follows the almost surely convergence

$$\Delta_n = \phi(Y_{[\gamma n]:n}^{(1)}, Y_{[\gamma n]:n}^{(2)}) \longrightarrow \phi(q_{\gamma}^{(1)}, q_{\gamma}^{(2)}) = \Delta \text{ as } n \rightarrow \infty.$$

Finally, under the premise that $\Delta < \delta$ is valid, it follows the almost surely convergence

$$\Delta_n - \delta = \Delta_n - \Delta + \Delta - \delta \longrightarrow \Delta - \delta < 0 \text{ as } n \rightarrow \infty$$

and so almost surely

$$T_n = \sqrt{n}(\Delta_n - \delta) \longrightarrow -\infty \text{ as } n \rightarrow \infty,$$

that is, (a). Under the premise that $\Delta > \delta$ is true, we have analogously the almost surely convergence

$$\Delta_n - \delta = \Delta_n - \Delta + \Delta - \delta \longrightarrow \Delta - \delta > 0 \text{ as } n \rightarrow \infty$$

and so the almost surely convergence

$$T_n = \sqrt{n}(\Delta_n - \delta) \longrightarrow +\infty \text{ as } n \rightarrow \infty,$$

that is, (c).

For the proof of (b), we use empirical process theory and apply the (functional) Delta method. For this purpose, it is useful to regard distribution functions defined on \mathbb{R} or \mathbb{R}^2 also as functions on $\overline{\mathbb{R}} = [-\infty, \infty]$ or $\overline{\mathbb{R}}^2 = [-\infty, \infty]^2$. In fact, $\overline{\mathbb{R}}^2$ can be endowed with the metric

$$d(x, y, u, v) = \max(|\arctan x - \arctan u|, |\arctan y - \arctan v|), \quad (x, y), (u, v) \in \overline{\mathbb{R}}^2,$$

such that $(\overline{\mathbb{R}}^2, d)$ is a compact metric space. The empirical forecast distribution function of the own funds in model run $k = 1, 2$ is given by

$$F_n^{(k)}(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}(Y_j^{(k)} \leq x), \quad x \in \overline{\mathbb{R}},$$

that is an estimator for the related modeled forecast distribution function of the own funds

$$F^{(k)}(x) = \mathbb{P}(Y^{(k)} \leq x), \quad x \in \overline{\mathbb{R}}.$$

The bivariate empirical forecast distribution function of the own funds of model runs 1 and 2 is

$$F_n(x, y) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}(Y_j^{(1)} \leq x, Y_j^{(2)} \leq y), \quad (x, y) \in \overline{\mathbb{R}}^2,$$

that is an estimator for the bivariate forecast distribution function of the own funds of model runs 1 and 2

$$F(x, y) = \mathbb{P}(Y^{(1)} \leq x, Y^{(2)} \leq y), \quad (x, y) \in \overline{\mathbb{R}}^2.$$

The related empirical process is given by

$$\sqrt{n}(F_n(x, y) - F(x, y)), \quad (x, y) \in \overline{\mathbb{R}}^2.$$

Note that

$$F_n(x, \infty) = F_n^{(1)}(x), \quad x \in \overline{\mathbb{R}}, \quad F_n(\infty, x) = F_n^{(2)}(x), \quad x \in \overline{\mathbb{R}},$$

and

$$F(x, \infty) = F^{(1)}(x), \quad x \in \overline{\mathbb{R}}, \quad F(\infty, x) = F^{(2)}(x), \quad x \in \overline{\mathbb{R}}.$$

Denote by U_n a stochastic process on a probability space $(\Omega, \mathfrak{A}, P)$ with sample paths in

$$\ell^\infty(\overline{\mathbb{R}}^2) = \{f : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}; \|f\|_{\overline{\mathbb{R}}^2} < \infty\},$$

where $\|f\|_{\overline{\mathbb{R}}^2} = \sup_{(x,y) \in \overline{\mathbb{R}}^2} |f(x, y)|$, $f : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}$, is the uniform norm. The map $(f_1, f_2) \mapsto \|f_1 - f_2\|_{\overline{\mathbb{R}}^2}$, $(f_1, f_2) \in \ell^\infty(\overline{\mathbb{R}}^2) \times \ell^\infty(\overline{\mathbb{R}}^2)$, defines a metric on $\ell^\infty(\overline{\mathbb{R}}^2)$. Denote by $\mathfrak{B}(\ell^\infty(\overline{\mathbb{R}}^2))$

the Borel σ -field on $(\ell^\infty(\overline{\mathbb{R}}^2), \|\cdot\|_{\overline{\mathbb{R}}^2})$. Let U be another stochastic process, for simplicity defined on the same probability space, with sample paths in $\ell^\infty(\overline{\mathbb{R}}^2)$. It is

$$U_b(\overline{\mathbb{R}}^2, d) = \{f \in \ell^\infty(\overline{\mathbb{R}}^2); f \text{ is uniformly } d\text{-continuous}\}$$

a separable subspace of $\ell^\infty(\overline{\mathbb{R}}^2)$. If the process U has sample paths in $U_b(\overline{\mathbb{R}}^2, d)$, the process U is $(\mathfrak{A}, \mathfrak{B}(\ell^\infty(\overline{\mathbb{R}}^2)))$ -measurable. We say that U_n converges in distribution to U , in notation $U_n \xrightarrow{d} U$ as $n \rightarrow \infty$, if U is $(\mathfrak{A}, \mathfrak{B}(\ell^\infty(\overline{\mathbb{R}}^2)))$ -measurable and

$$\forall f: \ell^\infty(\overline{\mathbb{R}}^2) \rightarrow \mathbb{R} \text{ continuous and bounded: } \lim_{n \rightarrow \infty} E^*(f(U_n)) = E(f(U)),$$

where E^* is the outer expectation, see, for example, van der Vaart & Wellner (1996) for details. Combining the results of Dudley (1984) with Lemma 2.6.17 (i) and (ii) in van der Vaart & Wellner (1996), we have that

$$C = \{(u, v) \in \mathbb{R}^2; u \leq x, v \leq y\}; (x, y) \in \overline{\mathbb{R}}^2\}$$

is a Vapnik–Chervonenkis class. Then, the set

$$\mathcal{V} = \{f: \mathbb{R}^2 \rightarrow \mathbb{R}, f(u, v) = I((u, v) \in C), (u, v) \in \overline{\mathbb{R}}^2, C \in \mathcal{C}\}$$

is a Vapnik–Chervonenkis graph class. The empirical process $\sqrt{n}(F_n - F)$ has sample paths in $\ell^\infty(\overline{\mathbb{R}}^2)$ and is indexed by the Vapnik–Chervonenkis graph class. We apply the general convergence result 4.2 in Ziegler (1997) to our empirical process. It is clear that \mathcal{V} can be equipped with a metric corresponding to d such that the related space is totally bounded. From Lemma 2.4 in Ziegler (1997), it follows that \mathcal{V} has uniformly integrable entropy. From Lindeberg’s condition combined with the Cramér–Wold device the convergence in distribution of the finite dimensional projections of our empirical process to centered multivariate normal distributions follows. With the envelope $g(u, v) = 1, (u, v) \in \mathbb{R}^2$, the conditions (8) and (10) in 4.2 in Ziegler (1997) are directly satisfied. From our assumption that F is uniformly continuous we deduce that condition (9) in 4.2 in Ziegler (1997) holds. In all, we obtain from 4.2 in Ziegler (1997) the convergence in distribution

$$\sqrt{n}(F_n - F) \xrightarrow{d} G \text{ as } n \rightarrow \infty,$$

where G is a centered Gaussian process with sample paths in $U_b(\overline{\mathbb{R}}^2, d)$ and covariance function

$$c(x, y, u, v) = F(x \wedge u, y \wedge v) - F(x, y)F(u, v), (x, y), (u, v) \in \overline{\mathbb{R}}^2.$$

Noticing that in a finite-dimensional space Hadamard differentiability is just the classical total differentiability, the map $\phi: \mathbb{R}^2 \rightarrow (0, \infty)$ introduced above is Hadamard differentiable in $(q_\gamma^{(1)}, q_\gamma^{(2)})$ tangentially to \mathbb{R}^2 with derivative

$$\begin{aligned} \phi'_{(q_\gamma^{(1)}, q_\gamma^{(2)})}(x, y) = & \frac{x \cdot \text{SCR}^{(2)} + y \cdot \text{SCR}^{(1)}}{(\text{SCR}^{(1)})^2} I\left(\frac{\text{SCR}^{(2)} - \text{SCR}^{(1)}}{\text{SCR}^{(1)}} > 0\right) \\ & - \frac{x \cdot \text{SCR}^{(2)} + y \cdot \text{SCR}^{(1)}}{(\text{SCR}^{(1)})^2} I\left(\frac{\text{SCR}^{(2)} - \text{SCR}^{(1)}}{\text{SCR}^{(1)}} < 0\right), (x, y) \in \mathbb{R}^2. \end{aligned}$$

Denoting by \mathcal{F} the space of all functions $H : \overline{\mathbb{R}}^2 \rightarrow \mathbb{R}$, where for $H \in \mathcal{F}$ the functions $H^{(1)} = H(\cdot, \infty)$ and $H^{(2)} = H(\infty, \cdot)$ are càdlàg, which can be endowed with the usual supremum metric, we deduce from the results for the Hadamard differentiability of the quantile functional, see Theorem 21.3. in van der Vaart & Wellner (1996), that the functional $\varphi : \mathcal{F} \rightarrow \mathbb{R}^2$, defined by

$$\varphi(H) = \left((H^{(1)})^{-1}(\gamma), (H^{(2)})^{-1}(\gamma) \right), H \in \mathcal{F},$$

is Hadamard differentiable in F tangentially to \mathcal{H} , where $\mathcal{H} \subset \mathcal{F}$ is the subspace of \mathcal{F} of functions continuous in $(q_\gamma^{(1)}, q_\gamma^{(2)})$, with Hadamard derivative

$$\varphi'_F(H) = \left(-\frac{H^{(1)}(q_\gamma^{(1)})}{f^{(1)}(q_\gamma^{(1)})}, -\frac{H^{(2)}(q_\gamma^{(2)})}{f^{(2)}(q_\gamma^{(2)})} \right), H \in \mathcal{H}.$$

Defining the functional $\psi : \mathcal{F} \rightarrow \mathcal{Y}$ by

$$\psi(H) = \left| \frac{y^{(2)} - (H^{(2)})^{-1}(\gamma) - (y^{(1)} - (H^{(1)})^{-1}(\gamma))}{y^{(1)} - (H^{(1)})^{-1}(\gamma)} \right|, H \in \mathcal{F},$$

it follows from the chain rule, see Theorem 20.9 in van der Vaart (1998), that $\psi = \phi \circ \varphi$ is Hadamard differentiable in F tangentially to \mathcal{H} with Hadamard derivative $\psi'_F = \phi'_{(q_\gamma^{(1)}, q_\gamma^{(2)})} \circ \varphi'_F$, where

$$\begin{aligned} \psi'_F(H) = & \frac{-\frac{H^{(1)}(q_\gamma^{(1)})}{f^{(1)}(q_\gamma^{(1)})} \cdot \text{SCR}^{(2)} - \frac{H^{(2)}(q_\gamma^{(2)})}{f^{(2)}(q_\gamma^{(2)})} \cdot \text{SCR}^{(1)}}{(\text{SCR}^{(1)})^2} I\left(\frac{\text{SCR}^{(2)} - \text{SCR}^{(1)}}{\text{SCR}^{(1)}} > 0\right) \\ & - \frac{-\frac{H^{(1)}(q_\gamma^{(1)})}{f^{(1)}(q_\gamma^{(1)})} \cdot \text{SCR}^{(2)} - \frac{H^{(2)}(q_\gamma^{(2)})}{f^{(2)}(q_\gamma^{(2)})} \cdot \text{SCR}^{(1)}}{(\text{SCR}^{(1)})^2} I\left(\frac{\text{SCR}^{(2)} - \text{SCR}^{(1)}}{\text{SCR}^{(1)}} < 0\right), H \in \mathcal{H}. \end{aligned}$$

Under $\Delta = \delta$, it follows from the (functional) Delta method, see Theorem 20.8 in van der Vaart (1998), that

$$T_n = \sqrt{n}(\Delta_n - \Delta) = \sqrt{n}(\psi(F_n) - \psi(F)) \xrightarrow{d} \psi'_F(G) = N \text{ as } n \rightarrow \infty.$$

Because the Hadamard derivative ψ'_F is a linear and continuous functional and G is a centered Gaussian process with continuous paths, N has a centered normal distribution. By simple calculation we obtain the variance of the distribution of N as it is stated. In all, b) is proved. □

Finally, we prove of Theorem 2. Now, we use the fact that the empirical process theory in Ziegler (1997) is developed in a general setting for triangular arrays of random variables. In addition, a general Glivenko–Cantelli result in Gänbler & Ziegler (1994) is applied. Finally, the concept of Hadamard differentiability and the functional Delta method for bootstrap in van der Vaart (1998) yields the statement.

Proof. Proof of Theorem 2. To prove the statement we argue analogously to the Proof of Theorem 1 (b). In the bootstrap situation, we consider the empirical distribution functions

$$\hat{F}_n^{(k)}(x) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}(\hat{Y}_{j,n}^{(k)} \leq x), \quad x \in \overline{\mathbb{R}},$$

$k = 1, 2$, and

$$\hat{F}_n(x, y) = \frac{1}{n} \sum_{j=1}^n \mathbf{I}(\hat{Y}_{j,n}^{(1)} \leq x, \hat{Y}_{j,n}^{(2)} \leq y), \quad (x, y) \in \overline{\mathbb{R}}^2,$$

as well as the empirical process

$$\sqrt{n}(\hat{F}_n(x, y) - F_n(x, y)), \quad (x, y) \in \overline{\mathbb{R}}^2,$$

based on triangular arrays of row-wise independent and identically distributed random variables with underlying distributions $F_n^{(k)}$ and F_n . From our assumptions on F , the regularity conditions of the general convergence result in Gänbler & Ziegler (1994) can also be easily verified. At first, we obtain from Corollary 4.1 (ii) in Gänbler & Ziegler (1994) almost surely

$$F_n \longrightarrow F \text{ as } n \rightarrow \infty,$$

where the convergence holds in the uniformly sense. Noticing that 4.2 in Ziegler (1997) is valid for triangular arrays of random variables as well, the latter convergence result implies that the conditions in 4.2 in Ziegler (1997) hold almost surely analogous to the Proof of Theorem 1 (b). Therefore, it follows the conditionally convergence in distribution

$$\sqrt{n}(\hat{F}_n - F_n) \xrightarrow{d} G \text{ as } n \rightarrow \infty$$

given the original observations almost surely analogous to the Proof of Theorem 1 (b), where G is a centered Gaussian process with sample paths in $U_b(\overline{\mathbb{R}}^2, d)$. In particular, the distribution of G is the same distribution as in the limit in the Proof of Theorem 1 (b) if $\Delta = \delta$ is valid. Noticing that the convergence in distribution of a process with values in $\ell^\infty(\overline{\mathbb{R}}^2)$ to another tight process with values in $\ell^\infty(\overline{\mathbb{R}}^2)$ can be expressed with the help of Lipschitz continuous functions, compare with Chapter 23 in van der Vaart (1998), the stated conditionally convergence in distribution given the original observations almost surely can be formulated as the following almost surely convergence:

$$\sup_{h \in \text{BL}_1(\ell^\infty(\overline{\mathbb{R}}^2))} |E(h(\sqrt{n}(\hat{F}_n - F_n))|(Y_1^{(1)}, Y_1^{(2)}), \dots, (Y_n^{(1)}, Y_n^{(2)})) - E(h(G))| \longrightarrow 0 \text{ as } n \rightarrow \infty,$$

where $\text{BL}_1(\ell^\infty(\overline{\mathbb{R}}^2))$ denotes the set of all Lipschitz continuous functions $h: \ell^\infty(\overline{\mathbb{R}}^2) \rightarrow [-1, +1]$ with $|h(f) - h(g)| \leq \|f - g\|_{\overline{\mathbb{R}}^2}$ for all $f, g \in \ell^\infty(\overline{\mathbb{R}}^2)$. Note that the process G is tight because G has sample paths in $U_b(\overline{\mathbb{R}}^2, d)$. Because almost surely convergence implies convergence in probability, it follows the conditionally convergence in distribution

$$\sqrt{n}(\hat{F}_n - F_n) \xrightarrow{d} G \text{ as } n \rightarrow \infty$$

given the original observations in probability, where the stated conditionally convergence in distribution given the original observations in probability is in the sense of the following convergence in probability:

$$\sup_{h \in \text{BL}_1(\ell^\infty(\mathbb{R}^2))} |E(h(\sqrt{n}(\hat{F}_n - F_n))|(Y_1^{(1)}, Y_1^{(2)}, \dots, (Y_n^{(1)}, Y_n^{(2)})) - E(h(G))| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty,$$

see Chapter 23 in van der Vaart (1998). Because G is tight, and from the (functional) Delta method for bootstrap, see Theorem 23.9 in van der Vaart (1998), we obtain the conditionally convergence in distribution

$$\hat{T}_n = \sqrt{n}(\hat{\Delta}_n - \Delta_n) = \sqrt{n}(\psi(\hat{F}_n) - \psi(F_n)) \xrightarrow{d} \psi'_F(G) = N \text{ as } n \rightarrow \infty$$

given the original observations in probability, where N has a centered normal distribution, and the stated conditionally convergence in distribution given the original observations in probability is in the sense as it is stated below Theorem 2. In particular, the distribution of N is the same distribution as in the limit in Theorem 1 (b) if $\Delta = \delta$ is valid. □