Extending the Limits of Backtesting via the 'Vanishing p'-Approach

Yannick Hoga*

August 28, 2018

Abstract

We derive backtests of Value-at-Risk and Expected Shortfall forecasts for levels that vanish as a function of the sample size. In the standard case, the level of the forecasts is assumed to be fixed, leading to χ^2 -limiting distributions of the backtests. We show that for levels vanishing sufficiently fast, Poisson-type limits arise instead. These mimic key features of the test statistics, such as discreteness. Simulations demonstrate that for forecast levels and sample sizes of practical interest, using the Poisson-type limits leads to much improved size vis-à-vis the standard χ^2 -limits.

Keywords: Backtest, Expected Shortfall, Forecasting, Value-at-Risk JEL classification: C12 (Hypothesis Testing), C14 (Semiparametric and Nonparametric Methods), C52 (Model Evaluation, Validation, and Selection)

1 Motivation

Ample evidence for clustered volatility in financial time series has led to a burgeoning literature on dynamic forecasts of risk; e.g, Engle and Manganelli (2004), El Ghourabi, Francq, and Telmoudi (2016). Two of the most popular risk measures are the Value-at-Risk (VaR) and the Expected Shortfall (ES). The VaR at level p (VaR_{t,p}) is defined as the p-quantile of the conditional return distribution $F_t(x) = P \{Y_t \leq x \mid \Omega_{t-1}\}$, where Ω_{t-1} denotes the information available at time t - 1. Typically, the information set contains past returns, i.e., $\Omega_{t-1} = \sigma(Y_{t-1}, Y_{t-2}, ...)$. ES at level p (ES_{t,p}) is the conditional expected loss given a return below VaR_{t,p}, i.e., ES_{t,p} = $(1/p) \int_0^p \text{VaR}_{t,u} \, du$. To evaluate VaR and ES forecasts, the literature has proposed various so-called backtests; see, e.g., Escanciano and Olmo (2010) and Du and Escanciano (2017). All these tests assume a fixed level $p \in (0, 1)$ of the VaR and ES forecasts to be backtested. It is of interest for at least two reasons to extend backtests to the case where $p = p_n$ is allowed to converge to 0 as the sample size $n \to \infty$. First, some recent suggestions for producing VaR and ES forecasts based on extreme value theory explicitly let $p_n \to 0$ (Chan et al., 2007; Hoga, 2018+). Thus, it is desirable to develop corresponding backtests specifically

^{*}Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, yannick.hoga@vwl.uni-due.de. Support of DFG through grant HO 6305/1-1 is gratefully acknowledged.

tailored to vanishing p. Second, for test statistics of backtests, approximations of the finite-sample distributions suggested by standard asymptotic theory are often poor when the level p is small (Engle and Manganelli, 2004; Escanciano and Olmo, 2010). Escanciano and Olmo (2010, p. 44) argue that 'a different asymptotic theory based on $p \to 0$ as $n \to \infty$ may help to this end'.

It is the aim of this note to provide such a theory for the VaR backtest of Escanciano and Olmo (2010) and the ES backtest of Du and Escanciano (2017). For simplicity, we do so ignoring the issue of estimation risk that is dealt with in some detail by the aforementioned authors. The investigation of the above two backtests is only exemplary. So we consider this notes main contribution to be what may be termed the 'vanishing p'-approach (i.e., the approach of suitably letting $p = p_n \rightarrow 0$ in deriving limit theory for backtests) and highlighting its practical usefulness in simulations (in the sense of improved size).

2 'Vanishing p'-Backtests

Denote by Y_1, \ldots, Y_n the returns on a risky asset and let $p = p_n \in (0, 1/2)$, so that we focus on left-tail VaR and ES. To assess the quality of VaR forecasts $\widehat{\text{VaR}}_{t,p_n} = \widehat{F}_t^{\leftarrow}(p_n)$ (issued, e.g., from GARCH-type models) for these returns, we test

$$\mathcal{H}_0^{\mathrm{VaR}}: \qquad I_{t,n} := I_{\left\{Y_t \le \widehat{\mathrm{VaR}}_{t,p_n}\right\}} \overset{\mathrm{i.i.d.}}{\sim} \mathrm{Ber}(p_n),$$

where Ber(p) denotes the Bernoulli distribution with success probability p. So under the null, the sequence of VaR violations $\{I_{t,n}\}_{n\in\mathbb{N}, t=1,...,n}$ is a row-wise independent, identically distributed (i.i.d.) triangular array. Following Escanciano and Olmo (2010) and others, we consider the sample autoco-variances

$$\gamma_{j,n}^{I} = \frac{1}{n-j} \sum_{t=j+1}^{n} (I_{t,n} - p_n) (I_{t-j,n} - p_n), \qquad j \ge 1.$$
(1)

More precisely, $\gamma_{j,n}^{I}$ are the sample autocovariances calculated using the null hypothetical mean of the indicators, $E[I_{t,n} | \mathcal{H}_{0}^{VaR}] = p_n$, instead of the sample mean. This gives the test power against deviations from correct unconditional coverage, i.e., $E[I_{t,n}] = p_n$. A test of \mathcal{H}_{0}^{VaR} can then be based on the Ljung–Box-type test statistic

LB^{VaR}(d) =
$$n(n+2)\sum_{j=1}^{d} \frac{1}{n-j} \left(\frac{\gamma_{j,n}^{I}}{p_{n}(1-p_{n})}\right)^{2}, \quad d \ge 1.$$

Here, $p_n(1-p_n)$ in the denominator serves to standardize the indicators in (1) under the null, because $\operatorname{Var}(I_{t,n} - p_n \mid \mathcal{H}_0^{\operatorname{VaR}}) = p_n(1-p_n).$

We now turn to ES backtesting. To the best of our knowledge, the backtest of Du and Escanciano (2017) is the only ES backtest that is based on statistical theory (e.g., McNeil and Frey, 2000, base their ES backtest on a heuristic bootstrap procedure). To describe it, let the conditional distribution

 $F_t(\cdot) = P\{Y_t \leq \cdot \mid \Omega_{t-1}\}$ be continuous and, for the moment, known. The ES backtest is then based on the i.i.d. $\mathcal{U}[0, 1]$ -transformed residuals $U_t = F_t(Y_t)$, where $\mathcal{U}[0, 1]$ denotes the uniform distribution on [0, 1]. In practice, one uses the estimated version $\widehat{U}_t = \widehat{F}_t(Y_t)$.

Example 1. Consider a normal GARCH(1,1) model $Y_t = \sigma_t \varepsilon_t$ with i.i.d. standard normal innovations ε_t and $\sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2$. Since (under some mild conditions) σ_t^2 is measurable with respect to $\Omega_{t-1} = \sigma(Y_{t-1}, Y_{t-2}, \ldots)$, we have $F_t(\cdot) = \Phi(\cdot/\sigma_t)$ and thus $U_t = \Phi(Y_t/\sigma_t) = \Phi(\varepsilon_t)$, where $\Phi(\cdot)$ denotes the standard normal distribution function. The estimated counterparts are $\hat{F}_t(\cdot) = \Phi(\cdot/\hat{\sigma}_t)$, where $\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha} Y_{t-1}^2 + \hat{\beta} \hat{\sigma}_{t-1}^2$ for quasi-maximum likelihood (QML) estimates $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$, and $\hat{U}_t = \Phi(Y_t/\hat{\sigma}_t)$.

The key idea of Du and Escanciano (2017) is to write

$$\mathrm{ES}_{t,p_n} = E[Y_t \mid Y_t \leq \mathrm{VaR}_{t,p_n}, \ \Omega_{t-1}] = \frac{1}{p_n} \int_0^{p_n} \mathrm{VaR}_{t,u} \,\mathrm{d}u,$$

so that an ES backtest can be based on the *cumulative violations process* $H_{t,n}$ defined as

$$H_{t,n} = \frac{1}{p_n} \int_0^{p_n} I_{\{Y_t \le \widehat{\operatorname{VaR}}_{t,u}\}} \mathrm{d}u = \frac{1}{p_n} \int_0^{p_n} I_{\{\widehat{U}_t \le u\}} \mathrm{d}u = \frac{p_n - \widehat{U}_t}{p_n} I_{\{\widehat{U}_t \le p_n\}},$$

where we used that $I_{\{Y_t \leq \widehat{\operatorname{VaR}}_{t,u}\}} = I_{\{\widehat{U}_t \leq u\}}$. The cumulative violations process $H_{t,n}$ not only registers whether a VaR violation has occurred (via $I_{\{\widehat{U}_t \leq p_n\}} = I_{\{Y_t \leq \widehat{\operatorname{VaR}}_{t,p_n}\}}$), but also the magnitude of the violation (via $p_n - \widehat{U}_t = F_t(\operatorname{VaR}_{t,p_n}) - \widehat{F}_t(Y_t)$). Since, ideally, the estimated \widehat{U}_t mimic the i.i.d. $U_t \sim \mathcal{U}[0,1], \{H_{t,n}\}_{n \in \mathbb{N}, t=1,\dots,n}$ should ideally be row-wise i.i.d. with $H_{t,n} \stackrel{\mathcal{D}}{=} \frac{p_n - U}{p_n} I_{\{U \leq p_n\}}$ for $U \sim \mathcal{U}[0,1]$. This forms the basis for a test of

$$\mathcal{H}_0^{\mathrm{ES}}$$
: $H_{t,n}$ is i.i.d. with $H_{t,n} \stackrel{\mathcal{D}}{=} \frac{p_n - U}{p_n} I_{\{U \le p_n\}}, \quad U \sim \mathcal{U}[0,1].$

Similarly as above, we consider the sample autocovariances with the null hypothetical mean, $E[H_{t,n} | \mathcal{H}_0^{ES}] = p_n/2$, imposed:

$$\gamma_{j,n}^{H} = \frac{1}{n-j} \sum_{t=j+1}^{n} (H_{t,n} - p_n/2) (H_{t-j,n} - p_n/2), \qquad j \ge 1.$$

A test of $\mathcal{H}_0^{\text{ES}}$ can then be based on the Ljung–Box-type test statistic

LB^{ES}(d) =
$$n(n+2) \sum_{j=1}^{d} \frac{1}{n-j} \left(\frac{\gamma_{j,n}^{H}}{p_n(1/3 - p_n/4)} \right)^2$$
, $d \ge 1$.

As before, the term $p_n(1/3 - p_n/4)$ in the denominator serves as a standardization, because $\operatorname{Var}(H_{t,n} - p_n/2 \mid \mathcal{H}_0^{\mathrm{ES}}) = p_n(1/3 - p_n/4)$.

We can now state our main result.

Theorem 1. Let χ_d^2 denote a chi-squared distribution with d degrees of freedom and Poi(λ) a Poisson distribution with parameter λ .

(i) If $np_n^2 \to \infty$, then

$$\begin{split} \mathrm{LB}^{\mathrm{VaR}}(d) & \xrightarrow[(n \to \infty)]{\mathcal{D}} \chi_d^2 \qquad under \ \mathcal{H}_0^{\mathrm{VaR}}; \\ \mathrm{LB}^{\mathrm{ES}}(d) & \xrightarrow[(n \to \infty)]{\mathcal{D}} \chi_d^2 \qquad under \ \mathcal{H}_0^{\mathrm{ES}}. \end{split}$$

(ii) If $np_n^2 \to \lambda \in (0,\infty)$, then

$$\operatorname{LB}^{\operatorname{VaR}}(d) \xrightarrow[(n \to \infty)]{\mathcal{D}} Z^{\operatorname{VaR}}(d) \qquad under \ \mathcal{H}_0^{\operatorname{VaR}},$$

where $Z^{\text{VaR}}(d) = \lambda^{-1} \sum_{j=1}^{d} (Z_{\lambda}^{(j)} - \lambda)^2 \text{ for } Z_{\lambda}^{(j)} \stackrel{i.i.d.}{\sim} \text{Poi}(\lambda);$ $\text{LB}^{\text{ES}}(d) \xrightarrow[(n \to \infty)]{} Z^{\text{ES}}(d) \qquad under \mathcal{H}_0^{\text{ES}},$

where $Z^{\text{ES}}(d) = (9/\lambda) \sum_{j=1}^{d} (Z_{CP,\lambda}^{(j)} - \lambda/4)^2$ for i.i.d. $Z_{CP,\lambda}^{(j)}$ with compound Poisson distribution $\sum_{i=1}^{N} \widetilde{U}_i$, where $N \sim \text{Poi}(\lambda)$ and i.i.d. \widetilde{U}_i , distributed as the product U_1U_2 of two independent $\mathcal{U}[0,1]$ -random variables.

We provide a proof in the Appendix. The convergences in Theorem 1 are based on the idealized null hypotheses $\mathcal{H}_0^{\text{VaR}}$ and $\mathcal{H}_0^{\text{ES}}$. In the presence of estimation effects, these will not hold exactly and, thus, the convergences in Theorem 1 do not hold exactly. In this sense, we have not dealt with estimation effects. As pointed out above, Escanciano and Olmo (2010) and Du and Escanciano (2017) treat these in-depth.

Comparing Theorem 1 (i) with the fixed-p results of Escanciano and Olmo (2010) and Du and Escanciano (2017), we see that even when $p_n \to 0$ a χ^2 -limit may still be obtained. However, for this to be the case, p_n must converge to 0 at a slower rate than $1/\sqrt{n}$. In this sense, the requirement $np_n^2 \to \infty$ describes the barrier at which VaR and ES backtests can be applied as usual, i.e., with critical values calculated from the χ^2 -distribution. We refer to the other case where $np_n^2 \to \lambda$ as the 'vanishing p' case.

Comparing the cases $np_n^2 \to \infty$ and $np_n^2 \to \lambda$, we identify three marked differences. First, in the former case, standard Gaussian central limit theory applies, leading to a χ^2 -limit, i.e., a sum of i.i.d. standardized normal random variables. In contrast, in the latter case, a Poisson-type limit arises. More specifically, the limiting distribution is that of a sum of i.i.d. standardized Poisson random variables (for LB^{VaR}(d)) and a sum of i.i.d. standardized compound Poisson random variables (for LB^{ES}(d)). Second, no matter how fast or slow p_n may converge to zero, as long as $np_n^2 \to \infty$, a χ^2 -limit obtains. Once $np_n^2 \to \lambda$ holds, however, the speed of convergence of p_n to zero changes the limiting distribution via the parameter λ appearing in the Poisson-type limit. Third, while LB^{VaR}(d) and LB^{ES}(d) have the same same limit when $np_n^2 \to \infty$, this no longer holds under $np_n^2 \to \lambda$, where different limits arise.

Enlarging on these different limits, it is interesting to note that key features of the respective test



Figure 1: Quantile-quantile plot of χ_5^2 vs. LB^{VaR}(5) for $\lambda = 0.5$ and $\lambda = 2.5$ in (a) and (b), and χ_5^2 vs. LB^{ES}(5) for $\lambda = 0.5$ and $\lambda = 2.5$ in (c) and (d).

statistics are preserved in the asymptotic limit when $np_n^2 \to \lambda$. First, the test statistic LB^{VaR}(d) and its limit $Z^{\text{VaR}}(d)$ are discrete. A similar result also holds for LB^{ES}(d) and $Z^{\text{ES}}(d)$. As $H_{t,n}$ only has one atom in zero, $\gamma_{j,n}^H$ only has one atom in $p_n^2/4$. So LB^{ES}(d) and $Z^{\text{ES}}(d)$ both have continuous distributions outside the point $[(3/4)p_n/(1-3/4p_n)]^2n(n+2)\sum_{j=1}^d 1/(n-j)$ and its limit $(3/4)^2d\lambda$, respectively. This is illustrated in the quantile-quantile plot in Figure 1, where the χ_5^2 -distribution is compared to $Z^{\text{VaR}}(5)$ and $Z^{\text{ES}}(5)$ for different values of λ . Figure 1 shows that the larger λ , the closer LB^{VaR}(d) and LB^{ES}(d) are to the χ_5^2 -distribution. Yet, particularly for large quantiles that may be used as critical values in backtesting, notable differences still emerge. Second, the form of the test statistics is reflected in the limit, in that

$$\sum_{t=j+1}^{n} I_t I_{t-j} \xrightarrow[(n \to \infty)]{\mathcal{D}} Z_{\lambda}^{(j)} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} 1 \quad \text{and} \quad \sum_{t=j+1}^{n} H_t H_{t-j} \xrightarrow[(n \to \infty)]{\mathcal{D}} Z_{CP,\lambda}^{(j)} \stackrel{\mathcal{D}}{=} \sum_{i=1}^{N} \widetilde{U}_i,$$

as we show in the proof of Theorem 1. The summands 1 and \tilde{U}_i appearing in the limiting sums have a close connection to the key terms $I_t I_{t-j}$ and $H_t H_{t-j}$ appearing in $LB^{VaR}(d)$ and $LB^{ES}(d)$, viz.

 $1 \stackrel{\mathcal{D}}{=} I_t I_{t-j}|_{I_t=1, I_{t-j}=1} \quad \text{and} \quad \widetilde{U}_i \stackrel{\mathcal{D}}{=} H_t H_{t-j}|_{I_t=1, I_{t-j}=1}.$

(The first equality is trivial and the second equality can be checked readily by calculating the respective distribution functions.) The appearance of the $\operatorname{Poi}(\lambda)$ -distributed N as the upper limit in the sums can also be explained. Since $I_t I_{t-j} \sim \operatorname{Ber}(p_n^2)$, it is plausible that the number of successful Bernoulli trials $\sum_{t=j+1}^{n} I_t I_{t-j}$ is roughly binomial with parameters n and p_n^2 , written $\operatorname{Bin}(n, p_n^2)$. As is well-known, $\operatorname{Bin}(n, p_n^2) \approx \operatorname{Poi}(\lambda)$ for $np_n^2 \to \lambda$. So the limiting sum $\sum_{i=1}^{N} 1$ essentially sums up the terms $I_t I_{t-j}|_{I_t=1,I_{t-j}=1}$ as often as suggested by an approximation to the number of successful Bernoulli trials. A similar argument holds for $\sum_{i=1}^{N} \widetilde{U}_i$. Thus, the form of the test statistic is reflected in the limit. This close resemblance may explain the marked improvements gained over standard Gaussian asymptotics, that we demonstrate in simulations next.

3 Simulations

We explore the finite-sample properties of the tests derived from Theorem 1 for d = 5 and vanishing $p = p_n \to 0$ as $n \to \infty$. Specifically, we let $np_n^2 = 0.1, 0.3, \ldots, 2.9$. The value of λ is only specified asymptotically via $np_n^2 \to \lambda$. We simply set $\lambda = np_n^2$ in the simulations. As sample sizes, we use n = 250, 500, 1000 corresponding to about one year, two years and four years of daily stock returns. The simulation results are based on 1,000,000 replications.

To assess the impact of estimation effects on the backtests, we consider the normal GARCH(1,1) of Example 1 with parameters $(\omega, \alpha, \beta) = (0.00001, 0.1, 0.85)$ and QML estimates $(\widehat{\omega}, \widehat{\alpha}, \widehat{\beta})$, computed on a pre-sample Y_{-T+1}, \ldots, Y_0 of length $T \in \{1000, \infty\}$. The choice T = 1000 is motivated by Chan et al. (2007) and Hoga (2018+), and $T = \infty$ corresponds to the case of no estimation effect, i.e., knowledge of the true model parameters of the GARCH(1,1). The test of $\mathcal{H}_0^{\text{VaR}}$ is then based on $\{I_{\{Y_t \leq \widehat{\text{VaR}}_{t,p_n}\}}\}_{t=1,\ldots,n}$, where $\widehat{\text{VaR}}_{t,p_n} = \widehat{\sigma}_t \Phi^{-1}(p_n)$ with $\widehat{\sigma}_t^2 = \widehat{\omega} + \widehat{\alpha}Y_{t-1}^2 + \widehat{\beta}\sigma_{t-1}^2$. The test of $\mathcal{H}_0^{\text{ES}}$ is based on $\{\widehat{U}_t = \Phi(Y_t/\widehat{\sigma}_t)\}_{t=1,\ldots,n}$. So in both tests estimation effects appear, since σ_t needs to be estimated.

Since the limiting distribution $Z^{\text{VaR}}(d)$ for the VaR backtest is discrete, one cannot construct exact asymptotic level- α tests for any level $\alpha \in (0, 1)$. Hence, randomization is required. To circumvent this cumbersome procedure, we introduce a random variable $\varepsilon_0 \sim 0.001 \cdot N(0, 1)$ independent of $\{Y_t\}$ and exploit that under the $\mathcal{H}_0^{\text{VaR}}$,

$$\mathrm{LB}^{\mathrm{VaR}}(d) + \varepsilon_0 \xrightarrow[(n \to \infty)]{} Z^{\mathrm{VaR}}(d) + \varepsilon_0, \qquad \varepsilon_0 \text{ independent of } Z^{\mathrm{VaR}}(d).$$

The limiting distribution is now continuous, as desired. Note that for $LB^{ES}(d)$ no randomization is required, since the only point of discontinuity corresponds to a small quantile that is of no use as a critical value.

We consider three types of backtests. The first set uses the χ^2 -critical values for VaR and ES backtesting (from Theorem 1 (i)), the second uses critical values from $Z^{\text{VaR}}(d) + \varepsilon_0$ and $Z^{\text{ES}}(d)$ (from Theorem 1 (ii)), and the third derives critical values from the null hypothetical finite-sample distributions of $\text{LB}^{\text{VaR}}(d) + \varepsilon_0$ and $\text{LB}^{\text{ES}}(d)$, which can easily be simulated. This allows us to study—with (T = 1000) and without ($T = \infty$) estimation effects—how much closer our Poisson-type limits are to the finite-sample distribution compared to the standard χ^2 -limits. All critical values are calculated via simulations using 1,000,000 replications.

Figure 2 displays empirical size of the six backtests as a function of p_n . First, consider panels (a2), (b2) and (c2), where no estimation effects are present, i.e., $T = \infty$. The results for the third set of backtests are not displayed as they precisely hold size by construction in the absence of estimation risk. It is evident from the dotted lines that empirical size of the VaR and ES backtests with asymptotic χ^2 -critical values is rather poor for almost all small levels p_n considered here. This is in line with simulations by Escanciano and Olmo (2010). As they speculate, using limit theory in Theorem 1 (*ii*) tailored specifically to $p_n \to 0$ improves size. Even for the smallest sample size of n = 250, the VaR and ES backtests with Poisson-type limits hold size very well. For the ES backtest (hollow triangles), the improvements tend to be more marked. While the improvements in size are the most marked for small p_n , there is—maybe interestingly—no price to pay in terms of size for applying the 'vanishing p'-backtests for large p_n .

Turning to the case T = 1000 we find all tests to reject more frequently, presumably because the need to estimate parameters leads to more volatile VaR and ES forecasts, which are then rejected more often as inadequate. As demonstrated theoretically by Escanciano and Olmo (2010), the extent to which size distortions arise is determined by the ratio n/T. This is borne out in panels (a1), (b1) and (c1), where size distortions increase the larger n/T. A comparison of the χ^2 -type backtests with the other two reveals that the 'small p_n '-distortions dominate the estimation effects for more extreme levels. Nonetheless, even for small p_n , using the finite-sample distributions no longer leads to backtests that hold size and the 'vanishing p'-approach again leads to similar size, particularly for large n/T.

Compared with the χ^2 -limit theory, the 'vanishing p'-approach leads to backtests that have size close to the ones based on the finite-sample distributions, which are exact when no estimation effects are present. This suggests that the largest part of the size distortions when p_n is small can be eliminated by the 'vanishing p'-approach. The advantage of this approach over that exploiting finite-sample



Figure 2: Empirical size for VaR and ES backtests with χ^2 -critical values (χ^2 , dotted), critical values calculated from the Poisson-type limits $Z^{\text{VaR}}(d) + \varepsilon_0$ and $Z^{\text{ES}}(d)$ (Poi, solid), and critical values from $\text{LB}^{\text{VaR}}(d) + \varepsilon_0$ and $\text{LB}^{\text{ES}}(d)$ (Sim, dashed). Top, middle and bottom plots for n = 250,500 and 1000, respectively. Results shown for p_n with $np_n^2 = 0.1, 0.3, \ldots, 2.9$.

distributions of the test statistics is that incorporating estimation effects à la Escanciano and Olmo (2010) and Du and Escanciano (2017) is only possible in the (asymptotic) 'vanishing p'-approach. Figure 2 demonstrates that estimation effects can have serious consequences for size. Thus, it is well worth addressing jointly both distorting factors of backtests, viz. estimation effects and small p_n . Yet, this is beyond the scope of the present note.

We do not compare power for brevity and also because our key concern in this note was to improve size. It suffices to say that the oversized backtests based on χ^2 -critical values will trivially have higher power than the other tests. As documented by Du and Escanciano (2017), the ES backtests tend to have higher power than the VaR backtests, because they also take into account the magnitude of the violations.

4 Discussion

We show that two popular $\operatorname{VaR}_{t,p}$ and $\operatorname{ES}_{t,p}$ backtests can be applied as usual, when the level $p = p_n$ tends to zero not too fast as the sample size $n \to \infty$. The rate at which p_n is allowed to decay to zero is quite slow, which may explain the often observed poor size control of backtests when p_n is small. To address this, we derive limit theory for $np_n^2 \to \lambda \in (0, \infty)$. In this 'vanishing p'-approach, the limiting distributions depend on the decay rate of p_n via λ . The limit also reflects key properties of the test statistics. These advantages lead to marked improvements in size. Our theoretical results, derived for one VaR and one ES backtest, can of course only be exemplary. Yet, the finite-sample improvements afforded by the 'vanishing p'-approach suggest that if VaR/ES backtests are to be applied for small levels—for which they are of course intended—then these backtests should be developed for suitably vanishing p_n . Our simulations have shown that this holds particularly true for ES backtests, which are likely to gain in importance with the intended move of the Basel Committee on Banking Supervision (2013) from VaR to ES as the standard measure of risk. In future applications of the 'vanishing p'-approach it seems worthwhile to address also the size distortions induced by estimation effects.

References

- Basel Committee on Banking Supervision (2013). Fundamental Review of the Trading Book Second Consultative Document. Basel: Bank for International Settlements. URL: http://www.bis.org/publ/bcbs265.pdf.
- Chan, N. H. et al. (2007). "Interval Estimation of Value-at-Risk Based on GARCH Models with Heavy-Tailed Innovations". *Journal of Econometrics* 137, pp. 556–576.
- Du, Z. and J. C. Escanciano (2017). "Backtesting Expected Shortfall: Accounting for Tail Risk". Management Science 63, pp. 940–958.
- El Ghourabi, M., C. Francq, and F. Telmoudi (2016). "Consistent Estimation of the Value at Risk When the Error Distribution of the Volatility Model is Misspecified". *Journal of Time Series Analysis* 37, pp. 46–76.
- Engle, R. F. and S. Manganelli (2004). "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles". Journal of Business & Economic Statistics 22, pp. 367–381.
- Escanciano, J. C. and J. Olmo (2010). "Backtesting Parametric Value-at-Risk With Estimation Risk". Journal of Business & Economic Statistics 28, pp. 36–51.
- Hoga, Y. (2018+). "Confidence Intervals for Conditional Tail Risk Measures in ARMA-GARCH Models". Forthcoming in Journal of Business & Economic Statistics. DOI: 10.1080/07350015.2017.1401543. URL: http://www.tandfonline. com/doi/abs/10.1080/07350015.2017.1401543.
- McNeil, A. J. and R. Frey (2000). "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach". Journal of Empirical Finance 7, pp. 271–300.

Appendix

Proof of Theorem 1: We start by proving the results under (i), where $np_n^2 \to \infty$. First, consider LB^{VaR}(d) and define the $(d \times 1)$ -random vectors

$$\boldsymbol{y}_{t,n} = \frac{1}{p_n(1-p_n)} \left(\left[I_{t,n} - p_n \right] \left[I_{t-j,n} - p_n \right] \right)_{j=1,\dots,d}^{\top}, \qquad n \in \mathbb{N}, \ t = 1,\dots,n,$$

where we put $I_{0,n} = \ldots = I_{1-d,n} = p_n$. To derive that

$$n^{-1/2} \sum_{t=1}^{n} \boldsymbol{y}_{t,n} = \sqrt{n} \left(\frac{n-j}{n} \frac{\gamma_{j,n}}{p_n(1-p_n)} \right)_{j=1,\dots,d}^{\top} \xrightarrow{\mathcal{D}} N(\boldsymbol{0}, \boldsymbol{I}_d),$$
(A.1)

where I_d denotes the $(d \times d)$ -identity matrix and **0** a $(d \times 1)$ -vector of zeros, we apply a Cramér–Wold device and consider $\{a^{\top}y_{t,n}\}$ for a $(d \times 1)$ -vector a with $a^{\top}a = 1$. Set

$$\sigma_n^2 \coloneqq \operatorname{Var}\left(\sum_{t=1}^n \boldsymbol{a}^\top \boldsymbol{y}_{t,n}\right) = \sum_{t=1}^n \operatorname{Var}\left(\boldsymbol{a}^\top \boldsymbol{y}_{t,n}\right) = \sum_{t=1}^n \boldsymbol{a}^\top \operatorname{Var}\left(\boldsymbol{y}_{t,n}\right) \boldsymbol{a}.$$

Note for the above that the $y_{t,n}$ are serially uncorrelated. From $I_{t,n} \sim \text{Ber}(p_n)$ it is straightforward to derive that $\sigma_n^2/n \to 1$ as $n \to \infty$.

Since $\boldsymbol{a}^{\top}\boldsymbol{y}_{t,n}$ is (d+1)-dependent, we only need to verify the Lindeberg condition

$$\frac{1}{\sigma_n^2} \sum_{t=1}^n \mathrm{E}\left[(\boldsymbol{a}^\top \boldsymbol{y}_{t,n})^2 I_{\{|\boldsymbol{a}^\top \boldsymbol{y}_{t,n}| \ge \varepsilon \sigma_n\}} \right] \underset{(n \to \infty)}{\longrightarrow} 0 \qquad \forall \ \varepsilon > 0$$

to derive that $\sigma_n^{-1} \sum_{t=1}^n \boldsymbol{a}^\top \boldsymbol{y}_{t,n}$ is asymptotically standard normal (Utev, 1991, Corollary 2). Since $\sigma_n^2/n \to 1$, this implies

$$n^{-1/2} \sum_{t=1}^{n} \boldsymbol{a}^{\top} \boldsymbol{y}_{t,n} \xrightarrow[(n \to \infty)]{} N(0,1).$$
(A.2)

The indicator in the Lindeberg condition is eventually 0, since $\sigma_n p_n = (\sigma_n/\sqrt{n})(\sqrt{n}p_n) \to \infty$ by assumption on p_n . Thus, the Lindeberg condition is satisfied, so that (A.2) and hence (A.1) follow. Now, the convergence of LB^{VaR}(d) follows from (A.1) using the continuous mapping theorem and Slutzky's lemma.

The convergence of $LB^{ES}(d)$ can be derived similarly by considering the re-defined

$$\boldsymbol{y}_{t,n} = \frac{1}{p_n(1/3 - p_n/4)} \left(\left[H_{t,n} - p_n/2 \right] \left[H_{t-j,n} - p_n/2 \right] \right)_{j=1,\dots,d}^{\top}, \qquad n \in \mathbb{N}, \ t = 1,\dots,n,$$

where $H_{0,n} = \ldots = H_{1-d,n} = p_n/2$.

Now, we turn to verifying part (*ii*) of Theorem 1, where $np_n^2 \to \lambda$. We do so only for d = 2, because the proof for d > 2 is only notationally more complicated. Again, we first derive the convergence of $\text{LB}^{\text{VaR}}(d = 2)$. To do so, consider $\sum_{t=1}^{n} \tilde{y}_{t,n}$, where

$$\widetilde{\boldsymbol{y}}_{t,n} = ([I_t - p_n][I_{t-1} - p_n], [I_t - p_n][I_{t-2} - p_n])^{\top}$$

with $I_t := I_{t,n}$ for notational conciseness. Here, we require $\tilde{y}_{t,n}$ to be strictly stationary in rows for our limit theory. So for the moment we consider i.i.d. $\{I_t\}_{t=-1,0,\dots,n} \operatorname{Ber}(p_n)$ -random variables. Decompose

$$\sum_{t=1}^{n} \widetilde{y}_{t,n} = \sum_{t=1}^{n} \begin{pmatrix} I_t I_{t-1} \\ I_t I_{t-2} \end{pmatrix} - p_n \sum_{t=1}^{n} \begin{pmatrix} [I_t - p_n] + [I_{t-1} - p_n] \\ [I_t - p_n] + [I_{t-2} - p_n] \end{pmatrix} - \sum_{t=1}^{n} \begin{pmatrix} p_n^2 \\ p_n^2 \end{pmatrix} =: A - B - C.$$

We consider A, B and C separately. Since $np_n^2 \to \lambda$, we obtain $C \xrightarrow[(n \to \infty)]{} (\lambda, \lambda)^{\top}$. That $B = o_P(1)$ follows because for j = 0, 1, 2

$$P\left\{ \left| p_n \sum_{t=1}^n (I_{t-j} - p_n) \right| \ge \varepsilon \right\} \le \varepsilon^{-2} \operatorname{Var} \left\{ p_n \sum_{t=1}^n (I_{t-j} - p_n) \right\}$$
$$= (p_n/\varepsilon)^2 \sum_{t=1}^n \operatorname{Var}(I_{t-j})$$
$$= (p_n/\varepsilon)^2 n p_n (1 - p_n) = \mathcal{O}(n p_n^3) = o(1).$$

It remains to consider A. For brevity, define $\mathbf{z}_{t,n} = (I_t I_{t-1}, I_t I_{t-2})^{\top}$. We show that

$$A = \sum_{t=1}^{n} \boldsymbol{z}_{t,n} \xrightarrow[(n \to \infty)]{} (Z_{\lambda}^{(1)}, Z_{\lambda}^{(2)})^{\top}$$
(A.3)

for independent $Z_{\lambda}^{(i)} \sim \text{Poi}(\lambda)$ (i = 1, 2). To do so, we apply part (i_3) of Corollary 4.14 (ii) in Kobus (1995). Let $\|\cdot\|$ denote the maximum norm in \mathbb{R}^2 . We check the following conditions, which correspond to (4.23) and (4.9) in Kobus (1995),

$$\sup_{n \in \mathbb{N}} nP\left\{ \left\| \boldsymbol{z}_{1,n} \right\| > \varepsilon \right\} < M_{\varepsilon} < \infty;$$
(A.4)

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} nE \left[\left\| \boldsymbol{z}_{1,n} \right\| I_{\left\{ \left\| \boldsymbol{z}_{1,n} \right\| \le \delta \right\}} \right] = 0, \tag{A.5}$$

and, for $\boldsymbol{x} \in \mathbb{E}^6 = [\boldsymbol{0}, \boldsymbol{\infty}) \setminus \{\boldsymbol{0}\} \subset \mathbb{R}^6$,

$$nP\left\{(\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^{\top} \in [\boldsymbol{0}, \boldsymbol{x}]^{c}\right\} \xrightarrow[(n \to \infty)]{} \lambda \sum_{j=1}^{6} I_{\left\{\boldsymbol{x}_{j} \in [0,1]\right\}} \eqqcolon \nu^{\varphi}([\boldsymbol{0}, \boldsymbol{x}]^{c}),$$
(A.6)

which corresponds to (4.25) in Kobus (1995) by Resnick (2007, Lemma 6.1). The complement $[\mathbf{0}, \boldsymbol{x}]^c$ in (A.6) is taken in the space \mathbb{E}^6 .

Conditions (A.4) and (A.5) are easily verified, as

$$nP\left\{\left\|\boldsymbol{z}_{1,n}\right\| > \varepsilon\right\} \le n\sum_{j=1}^{2}P\left\{I_{1}I_{1-j} = 1\right\} = 2np_{n}^{2} \underset{(n \to \infty)}{\longrightarrow} 2\lambda$$

and, since $\|\boldsymbol{z}_{1,n}\|$ can only take on the values 0 and 1, we have $E[\|\boldsymbol{z}_{1,n}\| \| I_{\{\|\boldsymbol{z}_{1,n}\| \leq \delta\}}] = 0$ for $\delta < 1$. To derive (A.6), write $(\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^{\top} = (Z_1, \dots, Z_6)^{\top}$ and $\boldsymbol{x} = (x_1, \dots, x_6)^{\top}$. Then, use inclusion/exclusion to obtain

$$nP\left\{(\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^{\top} \in [\boldsymbol{0}, \boldsymbol{x}]^{c}\right\} = nP\left\{\bigcup_{j=1}^{6} \{Z_{j} > x_{j}\}\right\}$$
$$= \sum_{j=1}^{6} nP\left\{Z_{j} > x_{j}\right\} - \sum_{1 \le i < j \le 6} nP\left\{Z_{i} > x_{i}, Z_{j} > x_{j}\right\} + \dots (-1)^{6+1}nP\left\{\bigcap_{i=1}^{6} \{Z_{i} > x_{i}\}\right\}.$$

All terms but the first converge to zero. For the first, noting that all Z_j have the same distribution, we get

$$nP\left\{Z_j > x_j\right\} = nP\{I_1I_0 > x_j\}$$
$$= np_n^2 I_{\{x_j \in [0,1]\}} \xrightarrow[(n \to \infty)]{} \lambda I_{\{x_j \in [0,1]\}}.$$

Since for $Z_i > x_i$ and $Z_j > x_j$ $(i \neq j)$ to hold at least three distinct indicators I_t must equal one, which only occurs with probability $p_n^3 = P\{I_r = 1, I_s = 1, I_t = 1 \ (r, s, t \text{ distinct})\}$, the other terms are bounded by a finite number of terms of the form

$$nP\{Z_i > x_i, Z_j > x_j\} \le Knp_n^3 = o(1).$$

Combining the results gives (A.6), i.e.,

$$nP\left\{(\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^{\top} \in [\boldsymbol{0}, \boldsymbol{x}]^{c}\right\} \xrightarrow[(n \to \infty)]{} \lambda \sum_{j=1}^{6} I_{\{x_{j} \in [0,1]\}} = \nu^{\varphi}\left([\boldsymbol{0}, \boldsymbol{x}]^{c}\right).$$

Thus, ν^{φ} spreads mass onto each axis according to the one-dimensional measure $\nu([0, x]^c) := \lambda I_{\{x \in [0,1]\}}$ (i.e., it concentrates mass λ in the points $(1, 0, 0, 0, 0, 0), \ldots, (0, 0, 0, 0, 0, 1)$) but assigns no mass off the axes.

It remains to calculate $\rho(\cdot) := \nu_{x_0+x_1+x_2}^{\varphi}(\cdot) - \nu_{x_1+x_2}^{\varphi}(\cdot)$ defined in Equation (4.28) in Kobus (1995). With $\boldsymbol{y} = (y_1, y_2)^{\top} \in \mathbb{E}^2 = [\mathbf{0}, \boldsymbol{\infty}) \setminus \{\mathbf{0}\} \subset \mathbb{R}^2$

$$\begin{split} \nu_{x_0+x_1+x_2}^{\varphi} \left([\mathbf{0}, \boldsymbol{y}]^c \right) &:= \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 \ : \ (x_1, x_2)^\top + (x_3, x_4)^\top + (x_5, x_6)^\top \in [\mathbf{0}, \boldsymbol{y}]^c \} \right) \\ &= \nu^{\varphi} (\{ \boldsymbol{x} \in \mathbb{E}^6 \ : \ (x_1, 0)^\top \in [\mathbf{0}, \boldsymbol{y}]^c \}) + \nu^{\varphi} (\{ \boldsymbol{x} \in \mathbb{E}^6 \ : \ (0, x_2)^\top \in [\mathbf{0}, \boldsymbol{y}]^c \}) \\ &+ \dots + \nu^{\varphi} (\{ \boldsymbol{x} \in \mathbb{E}^6 \ : \ (x_5, 0)^\top \in [\mathbf{0}, \boldsymbol{y}]^c \}) + \nu^{\varphi} (\{ \boldsymbol{x} \in \mathbb{E}^6 \ : \ (0, x_6)^\top \in [\mathbf{0}, \boldsymbol{y}]^c \}) \\ &= \nu (\{ x \in (0, \infty) \ : \ x \in [0, y_1]^c \}) + \nu (\{ x \in (0, \infty) \ : \ x \in [0, y_2]^c \}) \\ &+ \dots + \nu (\{ x \in (0, \infty) \ : \ x \in [0, y_1]^c \}) + \nu (\{ x \in (0, \infty) \ : \ x \in [0, y_2]^c \}) \\ &= 3\lambda I_{\{ y_1 \in [0, 1]\}} + 3\lambda I_{\{ y_2 \in [0, 1]\}}, \end{split}$$

where we have used in the second step that ν^{φ} assigns no mass off the axes, such that if any component of $\boldsymbol{x} \in \mathbb{E}^6$ is not zero (e.g., $x_1 \neq 0$) then all other components can be set to zero ($x_2 = \ldots = x_6 = 0$). Similarly,

$$\nu_{x_1+x_2}^{\varphi}\left([\mathbf{0},\boldsymbol{y}]^c\right) \coloneqq \nu^{\varphi}\left(\{\boldsymbol{x}\in\mathbb{E}^6 : (x_3,x_4)^\top + (x_5,x_6)^\top\in[\mathbf{0},\boldsymbol{y}]^c\}\right)$$

$$= 2\lambda I_{\{y_1 \in [0,1]\}} + 2\lambda I_{\{y_2 \in [0,1]\}}$$

Consequently,

$$\rho([\mathbf{0}, \mathbf{y}]^c) = \lambda I_{\{y_1 \in [0,1]\}} + \lambda I_{\{y_2 \in [0,1]\}}.$$

Obviously, $\rho(\cdot)$ concentrates mass λ on the points $(1,0)^{\top}$ and $(0,1)^{\top}$. The conclusion in (A.3) follows from Corollary 4.14 of Kobus (1995).

Combining the convergences of A, B and C yields

$$\sum_{t=1}^{n} \widetilde{\boldsymbol{y}}_{t,n} \xrightarrow[(n \to \infty)]{\mathcal{D}} (Z_{\lambda}^{(1)} - \lambda, Z_{\lambda}^{(2)} - \lambda).$$

Setting $I_0 = I_{-1} = p_n$ does not change the convergence. To see this for the first component of $\tilde{y}_{t,n}$ (the second component can be dealt with similarly) write it as

$$\sum_{t=1}^{n} [I_t - p_n] [I_{t-1} - p_n] = [I_1 - p_n] [I_0 - p_n] + \sum_{t=2}^{n} [I_t - p_n] [I_{t-1} - p_n]$$

If we show that $[I_1 - p_n][I_0 - p_n] = o_P(1)$, the claim follows. Now, for n sufficiently large, s.t. $p_n < \varepsilon$,

$$P\{|[I_1 - p_n][I_0 - p_n]| \ge \varepsilon\} = P\{I_1 = 1, I_0 = 1\} = \mathcal{O}(p_n^2) = o(1)$$

With the so defined I_0 and I_1 ,

$$\sum_{t=1}^{n} \widetilde{\boldsymbol{y}}_{t,n} = ((n-1)\gamma_{1,n}^{I}, (n-2)\gamma_{2,n}^{I})^{\top} \xrightarrow[(n \to \infty)]{\mathcal{D}} (Z_{\lambda}^{(1)} - \lambda, Z_{\lambda}^{(2)} - \lambda)^{\top}.$$

From this, the continuous mapping theorem, Slutzky's lemma and $np_n^2 \to \lambda$,

$$LB^{VaR}(2) \xrightarrow[(n \to \infty)]{\mathcal{D}} \lambda^{-1} \left\{ (Z_{\lambda}^{(1)} - \lambda)^2 + (Z_{\lambda}^{(2)} - \lambda)^2 \right\}$$

This concludes the proof.

Finally, we turn to $LB^{ES}(d=2)$. The proof follows along similar lines as that for $LB^{VaR}(d=2)$, so we only sketch it using identical notation to highlight the similarities. Consider $\sum_{t=1}^{n} \tilde{y}_{t,n}$, with the re-defined

$$\widetilde{\boldsymbol{y}}_{t,n} = \left(\left[H_t - \frac{p_n}{2} \right] \left[H_{t-1} - \frac{p_n}{2} \right], \left[H_t - \frac{p_n}{2} \right] \left[H_{t-2} - \frac{p_n}{2} \right] \right)^\top,$$

where $H_t := H_{t,n}$. Consider i.i.d. $\{H_t\}_{t=-1,0,\dots,n}$. Decompose

$$\sum_{t=1}^{n} \widetilde{y}_{t,n} = \sum_{t=1}^{n} \begin{pmatrix} H_t H_{t-1} \\ H_t H_{t-2} \end{pmatrix} - \frac{p_n}{2} \sum_{t=1}^{n} \begin{pmatrix} [H_t - \frac{p_n}{2}] + [H_{t-1} - \frac{p_n}{2}] \\ [H_t - \frac{p_n}{2}] + [H_{t-2} - \frac{p_n}{2}] \end{pmatrix} - \frac{1}{4} \sum_{t=1}^{n} \begin{pmatrix} p_n^2 \\ p_n^2 \end{pmatrix} =: A - B - C.$$

We consider A, B and C separately. Similarly as before, we get $C \xrightarrow[(n \to \infty)]{} 1/4(\lambda, \lambda)^{\top}$ and $B = o_P(1)$.

It remains to consider A. Re-define $\mathbf{z}_{t,n} = (H_t H_{t-1}, H_t H_{t-2})^{\top}$. We show that

$$A = \sum_{t=1}^{n} \boldsymbol{z}_{t,n} \xrightarrow[(n \to \infty)]{} (Z_{CP,\lambda}^{(1)}, Z_{CP,\lambda}^{(2)})^{\top}$$
(A.7)

for independent $Z_{CP,\lambda}^{(i)}$ (i = 1, 2) with compound Poisson distribution $\sum_{i=1}^{N} \widetilde{U}_i$, where $N \sim \text{Poi}(\lambda)$ and i.i.d. \widetilde{U}_i , distributed as the product U_1U_2 of two independent $\mathcal{U}[0, 1]$ random variables.

To do so, we again apply part (i_3) of Corollary 4.14 (ii) in Kobus (1995). Conditions (A.4) and (A.5) are easy to establish. The equivalent of (A.6) is now

$$nP\left\{ (\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^{\top} \in [\boldsymbol{0}, \boldsymbol{x}]^{c} \right\} \xrightarrow[(n \to \infty)]{} \lambda \sum_{j=1}^{6} \left[1 - (x_{j} \land 1) \{ 1 - \log(x_{j} \land 1) \} \right] =: \nu^{\varphi}([\boldsymbol{0}, \boldsymbol{x}]^{c}) \quad (A.8)$$

for $\boldsymbol{x} \in \mathbb{E}^6$. Note that $P\{\widetilde{U}_1 > x\} = 1 - (x \wedge 1)\{1 - \log(x \wedge 1)\}$, as a simple calculation shows. To derive (A.8), write $(\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^\top = (Z_1, \dots, Z_6)^\top$ and $\boldsymbol{x} = (x_1, \dots, x_6)^\top$. Then, use inclusion/exclusion to obtain

$$nP\left\{(\boldsymbol{z}_{1,n}, \boldsymbol{z}_{2,n}, \boldsymbol{z}_{3,n})^{\top} \in [\boldsymbol{0}, \boldsymbol{x}]^{c}\right\} = nP\left\{\bigcup_{j=1}^{6} \{Z_{j} > x_{j}\}\right\}$$
$$= \sum_{j=1}^{6} nP\left\{Z_{j} > x_{j}\right\} - \sum_{1 \le i < j \le 6} nP\left\{Z_{i} > x_{i}, Z_{j} > x_{j}\right\} + \dots (-1)^{6+1}nP\left\{\bigcap_{i=1}^{6} \{Z_{i} > x_{i}\}\right\}.$$

By a similar argument as before, all terms but the first converge to zero. So we only consider a summand in the first term, where the Z_j again all have the same distribution. We have

$$nP\left\{Z_{j} > x_{j}\right\} = nP\{H_{1}H_{0} > x_{j}\}$$

$$= nP\left\{\frac{p_{n} - U_{1}}{p_{n}}I_{\{U_{1} \le p_{n}\}}\frac{p_{n} - U_{0}}{p_{n}}I_{\{U_{0} \le p_{n}\}} > x_{j}\right\}$$

$$= nE\left[I_{\left\{\frac{p_{n} - U_{1}}{p_{n}}I_{\{U_{1} \le p_{n}\}}\frac{p_{n} - U_{0}}{p_{n}}I_{\{U_{0} \le p_{n}\}} > x_{j}\right\}\right]$$

$$= n\int_{0}^{p_{n}}\int_{0}^{p_{n}}I_{\left\{\frac{p_{n} - u_{1}}{p_{n}}\frac{p_{n} - u_{0}}{p_{n}} > x_{j}\right\}}du_{1}du_{0}$$

$$= np_{n}^{2}[1 - (x_{j} \land 1)\{1 - \log(x_{j} \land 1)\}] \xrightarrow[(n \to \infty)]{}\lambda[1 - (x_{j} \land 1)\{1 - \log(x_{j} \land 1)\}].$$

Note $H_1H_0 \in [0,1]$. Hence, we obtain (A.8). The measure ν^{φ} in (A.8) spreads mass onto each axis according to the one-dimensional measure $\nu([0,x]^c) := \lambda[1 - (x \wedge 1)\{1 - \log(x \wedge 1)\}]$, but assigns no mass off the axes.

It remains to calculate
$$\rho(\cdot) := \nu_{x_0+x_1+x_2}^{\varphi}(\cdot) - \nu_{x_1+x_2}^{\varphi}(\cdot)$$
, where with $\boldsymbol{y} = (y_1, y_2)^{\top} \in \mathbb{E}^2$
 $\nu_{x_0+x_1+x_2}^{\varphi} \left([\boldsymbol{0}, \boldsymbol{y}]^c \right) := \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 : (x_1, x_2)^{\top} + (x_3, x_4)^{\top} + (x_5, x_6)^{\top} \in [\boldsymbol{0}, \boldsymbol{y}]^c \} \right)$
 $= \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 : (x_1, 0)^{\top} \in [\boldsymbol{0}, \boldsymbol{y}]^c \} \right) + \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 : (0, x_2)^{\top} \in [\boldsymbol{0}, \boldsymbol{y}]^c \} \right)$
 $+ \ldots + \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 : (x_5, 0)^{\top} \in [\boldsymbol{0}, \boldsymbol{y}]^c \} \right) + \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 : (0, x_6)^{\top} \in [\boldsymbol{0}, \boldsymbol{y}]^c \} \right)$

$$= \nu(\{x \in (0,\infty) : x \in [0,y_1]^c\}) + \nu(\{x \in (0,\infty) : x \in [0,y_2]^c\}) + \ldots + \nu(\{x \in (0,\infty) : x \in [0,y_1]^c\}) + \nu(\{x \in (0,\infty) : x \in [0,y_2]^c\}) = 3\lambda \left[1 - (y_1 \wedge 1)\{1 - \log(y_1 \wedge 1)\}\right] + 3\lambda \left[1 - (y_2 \wedge 1)\{1 - \log(y_2 \wedge 1)\}\right],$$

where we have used in the second step that ν^{φ} assigns no mass off the axes, such that if any component of $\boldsymbol{x} \in \mathbb{E}^6$ is not zero (e.g., $x_1 \neq 0$) then all other components can be set to zero ($x_2 = \ldots = x_6 = 0$). Similarly,

$$\nu_{x_1+x_2}^{\varphi} \left([\mathbf{0}, \boldsymbol{y}]^c \right) \coloneqq \nu^{\varphi} \left(\{ \boldsymbol{x} \in \mathbb{E}^6 : (x_3, x_4)^\top + (x_5, x_6)^\top \in [\mathbf{0}, \boldsymbol{y}]^c \} \right) \\ = 2\lambda \left[1 - (y_1 \wedge 1) \{ 1 - \log(y_1 \wedge 1) \} \right] \\ + 2\lambda \left[1 - (y_2 \wedge 1) \{ 1 - \log(y_2 \wedge 1) \} \right].$$

Consequently,

$$\rho([\mathbf{0}, \boldsymbol{y}]^c) = \lambda \left[1 - (y_1 \wedge 1) \{ 1 - \log(y_1 \wedge 1) \} \right] + \lambda \left[1 - (y_2 \wedge 1) \{ 1 - \log(y_2 \wedge 1) \} \right].$$

The convergence in (A.7) follows. Combining the convergences of A, B and C gives

$$\sum_{t=1}^{n} \widetilde{\boldsymbol{y}}_{t,n} \xrightarrow[(n \to \infty)]{\mathcal{D}} (Z_{CP,\lambda}^{(1)} - \lambda/4, Z_{CP,\lambda}^{(2)} - \lambda/4,)^{\top}$$

The remainder of the proof follows as before.

References

Kobus, M. (1995). "Generalized Poisson Distribution as Limits of Sums for Arrays of Dependent Random Vectors". Journal of Multivariate Analysis 52, pp. 199–244.

Resnick, S. (2007). Heavy-Tail Phenomena: Probabilistic and Statistical Modeling. New York: Springer.

Utev, S.A. (1991). "On the Central Limit Theorem for φ -Mixing Arrays of Random Variables". Theory of Probability and Its Applications 35, pp. 131–139.