Some differences in the rates of convergence of the sample covariance and covariation functions.

C. M. Gallagher* and T. Okuyama*

*Mathematical Sciences, Clemson University, Clemson, SC 29634, USA

Abstract

We compare the rate of convergence of the distribution functions of the sample covariance and covariation functions. We give a bound on the rate of convergence for the covariation for an independent identically distributed symmetric process in the domain of normal attraction of a non-normal or normal stable law. Since there is no known bound for the sample covariance in the non-normal case, we compare the rates using simulated data. If the process is symmetric α-stable, the sample covariation has an exact stable distribution, while the sample covariance appears to converge very slowly to its asymptotic distribution.

Keywords: Covariation; Covariance; Convergence rates;

1 Introduction

Let $\{Z_t\}$ be an independent identically distributed (*iid*) process with Z_1 symmetric and in the normal domain of attraction of an α-stable law of index $\alpha \in (1, 2]$ ($Z_1 \in \mathcal{ND}(\alpha)$), i.e.,

$$S_n = \sum_{t=1}^{n} n^{-1/\alpha} Z_t \Rightarrow S, \tag{1.1}$$

where S has a symmetric stable distribution with characteristic function

$$g_\alpha(t, \sigma) = e^{-|t\sigma|^\alpha}.$$

A sufficient condition for (1.1) is that

$$g_z(t) - 1 = -|\sigma t|^\alpha (1 + \mathrm{o}(1)) \quad \text{as} \quad \mathrm{t} \to 0,$$

where $g_z(t)$ denotes the characteristic function of Z_1. When $\alpha = 2$ this is implied by $EZ^2 < \infty$.

In this note we compare the rate of convergence of the distribution functions of the sample covariance and covariation functions for $\{Z_t\}$ described above. We will see that the sample covariation has an exact stable distribution when the data is symmetric stable and for non-stable symmetric data the rate of convergence is the same as that given in central limit theorems for *iid* data. Since no theoretical bound is available for summands in the sample covariance, we compare the convergence rates using synthetic data. Simulation evidence shows that for the covariance the rate of convergence is much slower. When the data is fat-tailed the sample correlation does not seem to have an improved rate of convergence.

If $\mathrm{EZ}_1^2 < \infty$ the covariance function at lag h,

$$\gamma(h) = \mathrm{E}(\mathrm{Z_t Z_{t-h}}),$$

is usually estimated with,

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{n-h} Z_t Z_{t+h}}{n-h} \quad \mathrm{h} = \quad 0, 1, 2, \ldots. \tag{1.2}$$

In this case $\sqrt{n}\hat{\gamma}(h)$ has an asymptotic normal distribution.

For $Z_1 \in \mathcal{ND}(\alpha)$ with $\alpha \in (1,2)$,

$$n\,(n\mathrm{ln}(\mathrm{n}))^{-1/\alpha}\,\hat{\gamma}(h) \Rightarrow U, \tag{1.3}$$

where U has a symmetric stable distribution with characteristic function,

$$g_\alpha(t) = E\mathrm{e}^{\mathrm{itU}} = \mathrm{e}^{\{-\mathrm{C}^2\Gamma(1-\alpha)\cos(\pi\alpha/2)|\mathrm{t}|^\alpha\}}, \tag{1.4}$$

where C is given by

$$x^\alpha P(Z_1 > x) \to C/2 \quad \text{as} \quad x \to \infty. \tag{1.5}$$

(See Davis and Resnick, 1986).

Runde (1997) and Phillips and Loretan (1990) both use the sample correlation

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)},$$

to develop test statistics to test for linear dependence in heavy tailed data. A problem with these tests is that the convergence under the *iid* hypothesis is quite slow. In fact, as the tail decay parameter α approaches 2, the rate of convergence appears to become increasingly slow. In some cases the Phillips and Loretan test may require a sample size on the order of 1 million in order to achieve a 5% significance level (See Adler et al., 1998).

A possible explanation for this slow distributional convergence is that the sample covariance magnifies the already fat tails. Even when the data is generated by a process in the normal domain of attraction of a stable law with $\alpha < 2$, the summands in the sample covariance (at any non-zero lag) do not belong to the normal domain of attraction of a stable law. Unlike the sample covariance, the sample covariation given below has the same tail behavior as Z_1.

Some recent papers have considered the sample covariation function,

$$\tau(h) = EZ_t\delta_{t-h},$$

where $\delta_t = \mathrm{sign}(\mathrm{Z_t})$, as a time series modeling tool. The sample covariation is given by,

$$\hat{\tau}(h,n) = \frac{\sum_{t=l}^{r} \delta_t Z_{t+h}}{n-h} \quad \mathrm{h} = \quad 0, \pm1, \pm2, \ldots \tag{1.6}$$

where $l = \max(1, 1+\mathrm{k})$ and $r = \min(\mathrm{n}, \mathrm{n}+\mathrm{k})$. Gallagher (2001A) considers using the covariation to estimate parameters in a symmetric stable p^{th} order autoregression AR(p). Simulation evidence in that paper indicates that for the AR(1) process 95% confidence intervals based on the covariation give better finite sample coverage than those using the sample ACF. Gallagher (1998) considers using the sample covariation as both an order identification tool for moving averages and as a diagnostic checking tool for residuals from a fitted ARMA model. A graphical procedure based on the covariation correctly identified the order of more simulated symmetric α-stable (normal) moving average time series than the usual procedure based on the sample ACF. Independence tests based on the sample covariation are proposed in Gallagher (2001B) and Gallagher (2001C). For simulated symmetric α-stable data these test statistics seem to converge faster to thier null limiting distribution than those which use the sample covariance. The purpose of this paper is in part to give an explanation for this.

The remainder of this paper is organized as follows. We consider bounds for the distributional convergence of the sample covariation in Section 2. In Section 3 we compare the convergence rates of the covariance and covariation at some specific quantiles using simulated data.

2 Convergence Rate for Covariation

In this section we investigate the rate of distributional convergence for the sample covariation. Theorem 2.1 below implies that

$$(n-h)^{1-1/\alpha}\hat{\tau}(h,n) \Rightarrow S,$$

at the same rate as S_n in (1.1). Although the theorem holds for any *iid* sequence of symmetric random variables which satisfy a central limit theorem, in this article we restrict our attention to sequences satisfying (1.1).

Let $G_n(x)$ be the distribution function of S_n and $G_\alpha(x,\sigma)$ be the symmetric α-stable distribution function. The rate of convergence of $G_n(x)$ to $G_\alpha(x,\sigma)$ is a well studied problem. For a nice overview in the infinite variance case, see Christoph (1991).

To state our main result let $F_n(x,h)$ be the distribution function of $(n-h)^{(1-1/\alpha)}\hat{\tau}(h,n)$. Theorem 2.1 is an immediate consequence Lemma 2.4.

Theorem 2.1 *Let $\{Z_t\}$ be an iid sequence of symmetric random variables in the normal domain of attraction of an α-stable law, then*

$$\sup_x |F_n(x,h) - G_\alpha(x,\sigma)| = \sup_x |G_{n-h}(x) - G_\alpha(x,\sigma)|.$$

Under assumptions on the relationship between the characteristic function of Z_1 and $g_\alpha(t,\sigma)$, such as (2.7) given below, explicit bounds on the convergence can be obtained. Condition (2.7) can be derived under moment or pseudo-moment assumptions (see Christoph, 1991).

Corollary 2.2 *If there exists an $\alpha < r \le 1+\alpha$, a constant C_r, and $\epsilon > 0$ such that for $|t| < \epsilon$*

$$|g_z(t) - g_\alpha(t,\sigma)| \le C_r |t|^r, \tag{2.7}$$

then

$$\sup_x |G_n(x) - G_\alpha(x,\sigma)| \le K_r n^{-(r-\alpha)/\alpha},$$

where K is a constant depending on r.

Remark 2.3 *If* $\mathrm{E}|\mathrm{Z}_1|^3 = \rho < \infty$ *then (2.7) is satisfied with* $r = 3$. *In this case the Berry-Essen inequality gives*

$$\sup_x |G_n(x\sqrt{2}\sigma) - \Phi(x)| \leq 3\rho(\sqrt{2}\sigma)^3 n^{-1/2},$$

where $E|Z_1|^2 = 2\sigma^2$.

Lemma 2.4 *Let* $\{Z_t\}$ *be an iid process. If* Z_1 *has a symmetric density, then for any integer* $h \neq 0$, *the process* $\{Z_t \delta_{t-h}\}$ *is an iid process with the same finite dimensional distributions as* $\{Z_t\}$.

Proof of Lemma 2.4. Without loss of generality we prove the result for $h = 1$. Using the fact that Z_2 is independent of δ_1 and the symmetry of Z_2,

$$\begin{aligned} P(Z_2\delta_1 < x_1) &= .5P(Z_2 < x_1) + .5P(Z_2 > -x_1) \\ &= P(Z_2 < x_1). \end{aligned}$$

Now assume,

$$P(Z_2\delta_1 < x_1, \ldots, Z_{n-1}\delta_{n-2} < x_{n-2}) = P(Z_2 < x_1, \ldots, Z_{n-1} < x_{n-2}), \tag{2.8}$$

then
$P(Z_2\delta_1 < x_1, \ldots, Z_n\delta_{n-1} < x_{n-1})$

$$\begin{aligned} &= P(Z_2\delta_1 < x_1, \ldots, Z_n < x_{n-1}, \delta_{n-1} = 1) \\ &+ P(Z_2\delta_1 < x_1, \ldots, Z_n > -x_{n-1}, \delta_{n-1} = -1) \\ &= P(Z_2\delta_1 < x_1, \ldots, Z_{n-1}\delta_{n-2} < x_{n-2}, \delta_{n-1} = 1)P(Z_n < x_{n-1}) \\ &+ P(Z_2\delta_1 < x_1, \ldots, Z_{n-1}\delta_{n-2} < x_{n-2}, \delta_{n-1} = -1)P(Z_n > -x_{n-1}) \\ &= P(Z_n < x_{n-1})[P(Z_2\delta_1 < x_1, \ldots, Z_{n-1}\delta_{n-2} < x_{n-2}, \delta_{n-1} = 1) \\ &+ P(Z_2\delta_1 < x_1, \ldots, Z_{n-1}\delta_{n-2} < x_{n-2}, \delta_{n-1} = -1)] \\ &= P(Z_n < x_{n-1})P(Z_2\delta_1 < x_1, \ldots, Z_{n-1}\delta_{n-2} < x_{n-2}). \end{aligned}$$

By assumption (2.8),

$$P(Z_2\delta_1 < x_1, \ldots, Z_n\delta_{n-1} < x_{n-1}) = P(Z_2 < x_1, \ldots, Z_n < x_{n-1}).$$

□

Remark 2.5 *If* Z_1 *has a symmetric* α*-stable density, then Lemma 2.4 implies that* $(n-h)^{1-1/\alpha}\hat{\tau}(h,n)$ *has the same distribution as* Z_1.

Remark 2.6 *If* $\{Y_t\}$ *is* iid *but not symmetric, the sequence given by*

$$Z_t = Y_{2t} - Y_{2t-1},$$

is iid *and symmetric.*

3 Simulations

In this section we will compare the rates of convergence on simulated data. Since statisticians are mostly concerned with the rate of convergence at particular quantiles of the distribution, we investigated the rate of convergence at the 0.95 and 0.975 quantiles of the limit distribution, corresponding to type I errors of 0.05 and .10 in a two-sided hypothesis test. These quantiles were chosen because of their common use in statistical hypothesis tests. In the infinite variance case we simulate from the symmetric stable density and a symmetric density with Pareto tails. In the finite variance case we considered the normal distribution and the t-distribution with 3 degrees of freedom. For all densities considered the simulations indicate that the distribution of the sample covariation converges faster at the above quantiles than does that of the sample covariance.

3.1 Infinite variance data

In the infinite variance case we compare the convergence rates using the symmetric stable density and a symmetric density with Pareto tails. As mentioned above, for symmetric stable data the sample covariation has an exact stable distribution. It has been shown that the central limit convergence for the Pareto distribution is quite slow $O(n^{(2-\alpha)/\alpha})$, so we expect slow convergence for the covariation.

We simulated from the symmetric Pareto density satisfying

$$P(Z_1 > x) = \frac{(x+1)^{-\alpha}}{2},$$

for $x > 0$. For this density both statistics have the same limiting distribution with characteristic function given by (1.4) with $C = 1$.

For the symmetric stable density we took scale $\sigma = 1$, corresponding to

$$C = \frac{1}{\Gamma(1-\alpha)\cos(\pi\alpha/2)}.$$

The limiting characteristic function for the normalized sample covariance is given by (1.4). Lemma 2.4 implies that for any n the covariation will have a stable distribution with $\sigma = 1$.

For $\alpha = 1.5, 1.7$, and 1.9 and each chosen sample size $n \leq 100,000$, we simulated 10,000 data sets. For each data set we calculated the sample covariance and covariation at lag 1. We calculated the proportion of statistics which fell below the quantiles of their respective limiting distribution. That is, we estimate the finite sample distribution function at the appropriate quantile. We used the symmetric stable quantiles given in Sammordnitsky and Taqqu (1990).

For stable data the convergence for the sample covariance seems to be fastest at the 0.975 quantile of the limiting distribution. Typical results for stable data are shown in Figure 1 and Figure 2 corresponding to $\alpha = 1.7$ and $\alpha = 1.9$, respectively. We see that even for $n = 100,000$ the estimated probability falls outside the 95% bounds $.975 \pm 1.96\sqrt{0.025 * 0.975/10000}$. At other quantiles (0.95 and 0.90) we get very similar graphs. The convergence for the lag 1 sample covariance slows as $\alpha \to 2$.

For Pareto data the convergence for both the covariation and covariance is very slow, especially as $\alpha \to 2$. However, our simulations indicate that the covariation is converging faster than the covariance. This can be seen in Figure 3 which gives simulation results for the 0.95 quantile when $\alpha = 1.7$. For $\alpha = 1.9$ neither statistic gives estimated probabilities within the approximate 95% bounds for any quantile considered even when $n = 100,000$.

Table 1: Number of esimated probabilities within the 95% bounds.

n	20000	25000	30000	35000	40000	45000	50000
γ	12	15	16	14	16	15	19
τ	19	18	19	17	19	19	19

3.2 Finite variance data

Here we simulate from the t and normal densities and compare the relative convergence of the empirical distribution functions at the 0.95 and 0.975 quantiles of the limiting normal distribution.

For the normal density, for all quantiles considered both statistics are well within the approximate 95% confidence bounds for all $n \geq 25$. So we do not include these simulation results here. In this case it seems the convergence of the sampling distribution of $\hat{\gamma}(1)$ is very fast and there is no apparent advantage to using the covariation - despite the fact that the sample covariation has an exact normal distribution in this case.

Figure 4 contains the empirical results for the sampling distribution of $\hat{\tau}(1)$, $\hat{\gamma}(1)$ and the sample correlation $\hat{\rho}(1)$ for the fat-tailed t-distribution. We see that the convergence of the sample covariance is again slower than the sample covariation. There seems to be little or no improvement for the sample correlation. To further assess the convergence for the t density, we ran 20 simulations for each n in increments of 5000 starting at $n = 30000$ (i.e., $n = 30000,\ 35000,\ 40000,\ \cdots$) until at least 19 of the estimates were found to be in the 95 % bounds. Table 1 shows the results of the simulations. We see that for $\hat{\gamma}(1)$ we get the expected 19 only when the sample size is 50,000.

4 Summary

Both the sample covariance and sample covariation are available as time series model fitting tools. The covariance has proved to be an excellent tool when the data is generated by a process with higher order moments and can still be used in the infinite variance case. However, as the tails of the underlying density become fat the convergence of the sampling distribution of $\hat{\gamma}(k)$ can become very slow. It may be very useful to have a finite sample correction for the sample covariance in the fat-tailed case. For an *iid* sequence of random variables with infinite second moment, there is currently no known bound on the rate of distributional convergence. In fact, even for *iid* summands in the non-normal domain of attraction of a stable law there is no Berry-Essen type inequality available.

For an *iid* sequence $\{Z_t\}$ of symmetric random variables which satisfies a central limit theorem the sample covariation converges at the same rate as the properly normalized sums $S_n = \sum_{t=1}^{n} Z_t$. For fat-tailed sequences the sample covariation seems to have an improved rate of convergence over the covariance at the quantiles commonly used by statisticians to test hypotheses and form confidence intervals. When Z_1 has an exact stable distribution then the covariation has an exact stable distribution. However, the covariation can still have a very slow convergence when the convergence in (1.1) is slow.

References

Adler, R. J., Feldman, R. E., and Gallagher, C., 1998. Analysing stable time series. In: Adler, R.J., Feldman, R.E. and M. S Taqqu, M.S. (Eds), A Practical Guide to Heavy Tails. Birkhäuser, Boston.

Christoph, G. (1991), On some differences in limit theorems with a normal and nonormal stable limit law, *Math. Nachr.* **153**, 247–256.

Davis, R. and Resnick, S. (1986), Limit theory for the sample covariance and correlation functions of moving averages, *Ann. Statist.* **14**, 533–558.

Gallagher, C., 1998. Fitting ARMA models to heavy-tailed data, University of California, Santa Barbara. PhD dissertation.

Gallagher, C., 2001A. A method for fitting stable autoregressive models using the covariation. Statistics and Probability Letters. To appear.

Gallagher, C., 2001B., A large sample independence test for finite mean data. Submitted for publication.

Gallagher, C., 2001C. A test for first order linear dependence in heavy-tailed data. Submitted for publication.

Phillips, P. C. B. and Loretan, M., 1991. The Durbin-Watson ratio under infinite variance errors. J. of Econometrics 47, 85-114.

Runde, R., 1998. The asymptotic null distribution of the Box-Pierce Q-statistic for random variables with infinite variance. J. of Econometrics 78, 205-216.

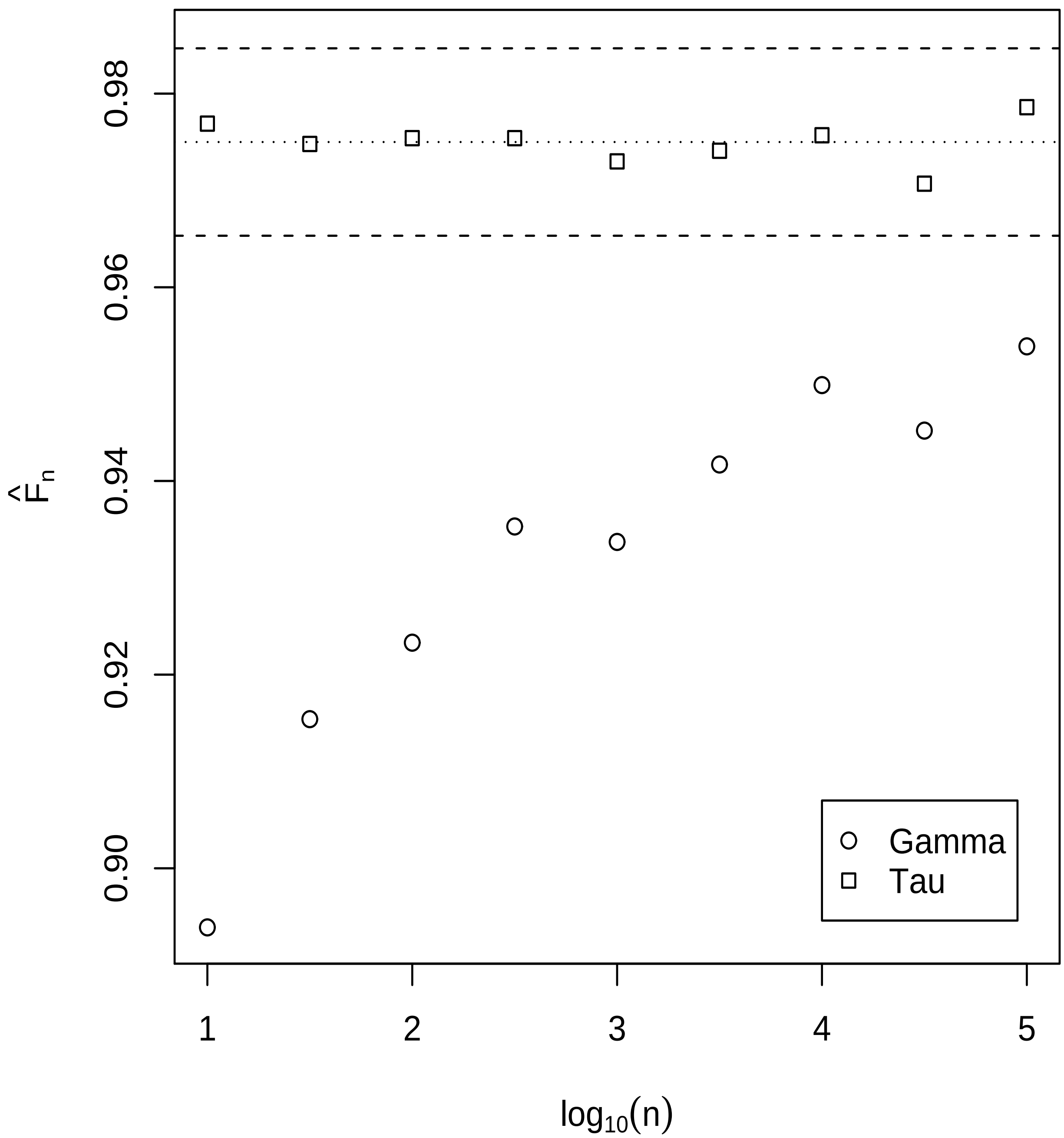

Figure 1: Relative convergence of the empirical sampling distribution of $\hat{\tau}(1)$ and $\hat{\gamma}(1)$ at the 0.975 quantile for symmetric stable data with $\alpha = 1.7$.

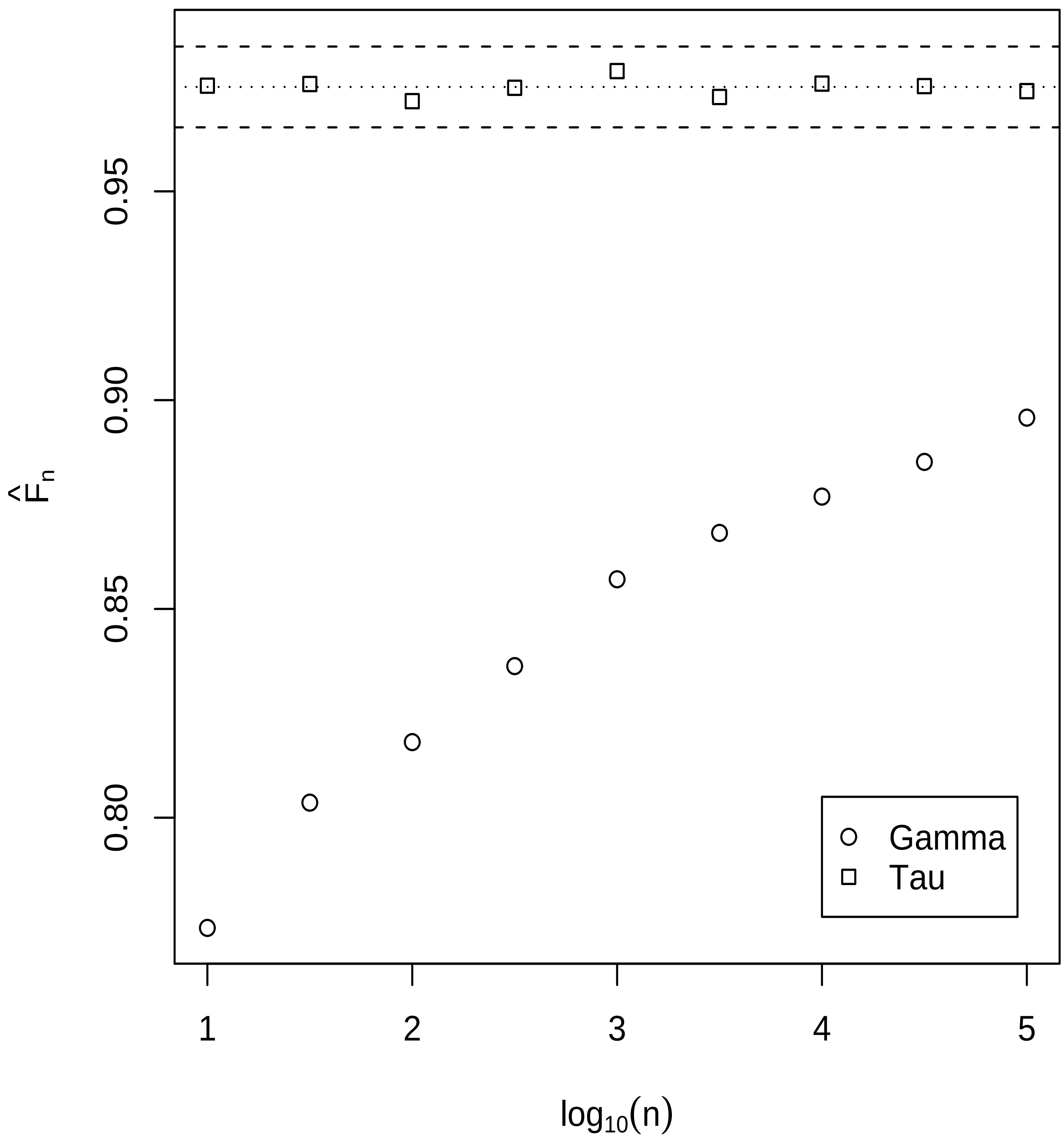

Figure 2: Relative convergence of the empirical sampling distribution of $\hat{\tau}(1)$ and $\hat{\gamma}(1)$ at the 0.975 quantile for symmetric stable data with $\alpha = 1.9$.

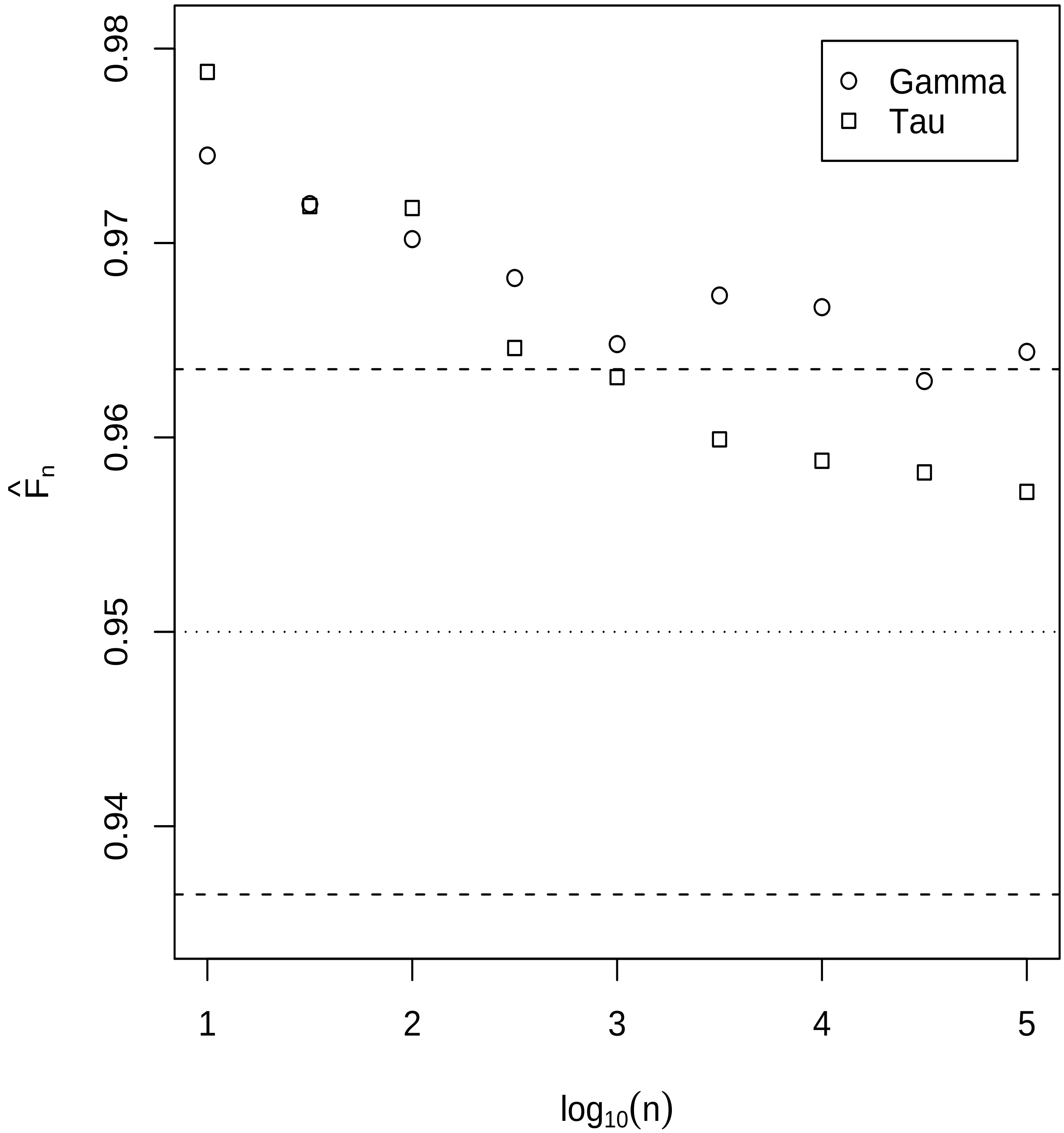

Figure 3: Relative convergence of the empirical sampling distribution of $\hat{\tau}(1)$ and $\hat{\gamma}(1)$ at the 0.95 quantile for symmetric Pareto data with $\alpha = 1.7$.

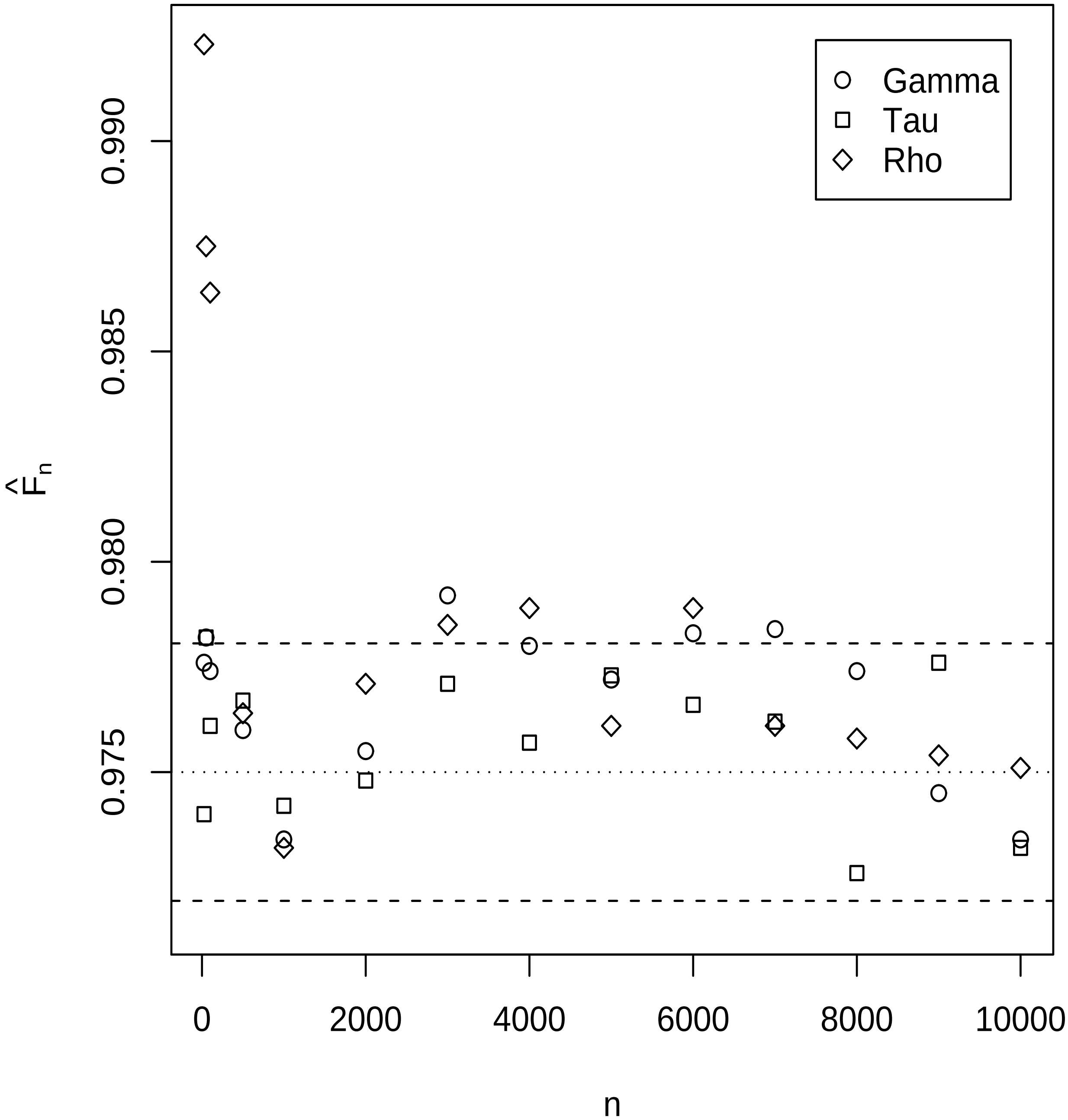

Figure 4: Convergence of the empirical sampling distribution of $\hat{\tau}(1)$, $\hat{\gamma}(1)$ and $\hat{\rho}(1)$ at the 0.975 quantile of the normal limiting distribution (1.96) for the t-distribution with 3 degrees of freedom.