Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Monte Carlo Methods for Testing Statistical Hypotheses

Optimizations Days 2011

Bruno Rémillard, HEC Montréal

Collaborators: C. Genest, K. Ghoudi, N. Papageorgiou, O. Scaillet

Presentation plan

1. Motivation
2. Goodness-of-fit
3. Bootstrap
4. Parametric bootstrap and applications
5. Multipliers method and applications

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Many important financial applications require knowing the underlying (multivariate) parametric distribution or related semiparametric distribution functions:

- Insurance
- Option pricing using optimal hedging (Remillard et al., 2010)
- Replication of hedge funds alla Kat-Palaro (Papageorgiou et al., 2008)
- Pricing of multiname credit derivatives (CDOs, n-th to default swaps, etc.) (Berrada et al., 2006)

Figure 1:S&P 500 over the period 12/31/1989 to 12/31/2009.

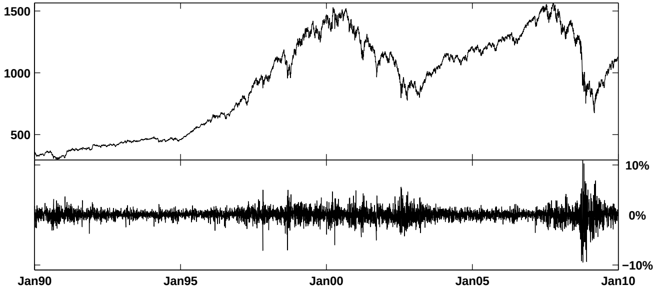


Table: Descriptive statistics for the S&P 500 daily returns.

Mean	Volatility	Skewness	Kurtosis
0.0002	0.0116	-0.1985	12.2536

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Traditional goodness-of-fit problem

To price options on the S&P500 index, one needs to know what is the underlying distribution of the returns.

One has to take into account the (possible) serial dependence, assuming stationarity and ergodicity.

The traditional goodness-of-fit problem is testing $H_0 : F \in \{F_\theta; \theta \in \mathcal{O}\}$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Functional estimations and related processes

To estimate F, one can use the nonparametric estimator $\hat{F}(x) = F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \leq x)$, $x \in \mathbb{R}^d$. The estimator is convergent, whatever H_0 is true or false.

Under H_0, if θ_n is an estimator of θ so that $\Theta_n = n^{1/2}(\theta_n - \theta) \rightsquigarrow \Theta$, F_{θ_n} is a consistent estimator of F. Therefore it makes sense to consider test statistics based on the empirical process $\mathbb{F}_n = \sqrt{n}\{F_n - F_{\theta_n}\}$.

When the data are serially independent, under H_0, $\mathbb{F}_n \rightsquigarrow \mathbb{F} = \mathbb{B}_{\mathbb{F}} - \Theta^\top \dot{F}$, where $\dot{F} = \nabla_{\theta_0} F$ and where $\mathbb{B}_F$ is a F-Brownian bridge, i.e., a continuous centered Gaussian process with covariance

$$\mathrm{Cov}(\mathbb{B}_F(x), \mathbb{B}_F(y)) = F(x \wedge y) - F(x)F(y), \quad x, y \in \mathbb{R}^d.$$

In general, the limiting distribution of $\mathbb{F} \neq \mathbb{B}_F$ depends on the the unknown parameters.

Goodness-of-fit test for normality

To illustrate the fact that estimating parameters affect the (asymptotic) distribution of the test statistic, consider doing a goodness-of-fit test of normality for the following simple model:

$$X_i = 1 + \varepsilon_i, \quad \varepsilon_i \sim N(0,1),$$

of independent observations.

For testing $H_0 : \varepsilon_i \sim N(0,1)$, one applies the Kolmogorov-Smirnov test, based on the statistic

$$\mathrm{KS}_n = \sup_{x\in\mathbb{R}} |\mathbb{F}_n(x)|.$$

The test statistic is evaluated with two sets of residuals:

$$e_{0i} = X_i - 1, \quad e_{1i} = X_i - \bar{X}, \quad i = 1, \ldots, n.$$

When taking residuals e_{0i}, $\mathbb{F}_n \rightsquigarrow \mathbb{F} = B_\Phi = \mathbb{B} \circ \Phi$, where $\mathbb{B}$ is the standard Brownian bridge, and Φ is the DF of a $N(0,1)$. Hence $\mathrm{KS}_n \rightsquigarrow \mathrm{KS} = \sup_{0 \le u \le 1} |\mathbb{B}(u)|$.

However, for residuals e_{1i}, $\mathbb{F}_n \rightsquigarrow \mathbb{F} = \mathbb{B} \circ \Phi - \mathcal{E}\phi$, where $n^{1/2}\bar{\varepsilon}_n \rightsquigarrow \mathcal{E}$, $\phi = \Phi'$. Hence,
$\mathrm{KS}_n \rightsquigarrow \sup_{0 \le u \le 1} |\mathbb{B}(u) - \mathcal{E}\phi \circ \Phi^{-1}(u)| \neq \mathrm{KS}$.

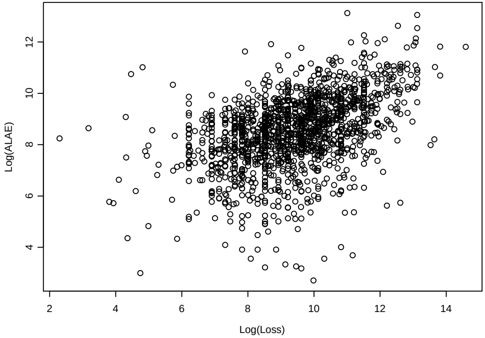
The results of the following simulation support that conclusion!

Table 1: Percentages of rejection of the standard Gaussian hypothesis for $N = 10000$ *replications of the Kolmogorov-Smirnov test, using samples of size* $n = 100$.

Residuals	
e_{0i}	e_{1i}
4.89%	.03%

Another example

Figure 2: Scatter plot of the natural logarithms of the indemnity payment (LOSS) and the allocated loss adjustmentexpense (ALAE) for 1500 general liability claims.



Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Question

For the data set presented in Figure 2, one is interested in finding a model of dependence, i.e., a copula, between the two variables.

A copula is just a multivariate distribution function with uniform marginals.

Recall that when the margins $F_1, \ldots, F_d$ are continuous, there is a unique copula C such that the joint distribution function F of X can be written as

$$F(x) = C\left\{F_1(x_1), \ldots, F_d(x_d)\right\},$$

$x = (x_1, \ldots, x_d) \in \mathbb{R}^d$.

So the question is: What is the underlying copula family? It is a (semiparametric) goodness-of-fit problem.

Goodness-of-fit problem for copulas

Suppose that independent observations $X_1, \ldots, X_n$ are drawn from law $F = C(F_1, \ldots, F_d)$ with (continuous) unknown marginals $F_1, \ldots, F_d$.

One is interested in testing $H_0 : C \in \mathcal{C} = \{C_\theta; \theta \in \mathcal{O}\}$.

Motivation
Traditional bootstrap
Parametric bootstrap
Multipliers
References

C can be estimated by the so-called empirical copula

$$C_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(e_{i,n} \le u), \qquad u \in [0,1]^d,$$

and $e_{ij,n} = \frac{\mathrm{Rank}(X_{ij})}{n}$, $j = 1, \ldots, d$, $1 \le i \le n$;

Also, under H_0, if θ_n is an estimator of θ so that $\Theta_n = n^{1/2}(\theta_n - \theta) \rightsquigarrow \Theta$, C_{θ_n} is a consistent estimator of C.

Therefore it makes sense to consider test statistics based on the empirical process $\mathbb{C}_n = \sqrt{n}\{C_n - C_{\theta_n}\}$.

However, $\mathbb{C}_n \rightsquigarrow \mathbb{C} = \mathbb{D} - \Theta^\top \dot{C}$. where $\dot{C} = \nabla_{\theta_0} C$, and $\mathbb{D}$ is the process that would be obtained if θ were known, which isn't the case! Furthermore, the law of $\mathbb{D}$ depends on C.

How can one determine P-values for statistics based on $\mathbb{C}_n$ in such a case?

Test of goodness-of-fit can also be based on Kendall's process (Genest et al., 2006) and Rosenblatt's transform (Genest et al., 2009).

Except a few exceptional cases, the law of the limiting processes depend on the unknown parameter θ.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

A first solution: Traditional bootstrap

In many cases, bootstrap can be used for testing $H_0 : F \in \{F_\theta; \theta \in \mathcal{O}\}$.

Given an estimation $\theta_n = T_n(X_1, \dots, X_n)$ of θ:

1. Generate, for $k = 1, \dots, N$, n independent observations $\hat{X}_{1,k}, \dots, \hat{X}_{n,k}$ from $X_1, \dots, X_n$.
2. Estimate θ by $\hat{\theta}^*_{k,n} = T_n\left(\hat{X}_{1,k}, \dots, \hat{X}_{n,k}\right)$.
3. Compute $\hat{\mathbb{F}}_{n,k} = \sqrt{n}\left(\hat{F}_{n,k} - F_{\hat{\theta}_{n,k}}\right) - \mathbb{F}_n$, where

$$\hat{F}_{n,k}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\hat{X}_{n,k} \leq x\right), \quad x \in \mathbb{R}^d.$$

The approximate P-value of a statistic $S_n = \phi_n(\mathbb{F}_n)$ is computed as

$$\frac{1}{N}\sum_{k=1}^{N} \mathbf{1}(S_{k,n}^{*} > S_n),$$

provided large value of S_n indicates lack-of-fit.

Works in general for $\mathbb{F}_n$ and $\mathbb{C}_n$ (Fermanian et al., 2004).

Another solution: Parametric bootstrap (PB)

PB for testing $H_0 : F \in \{F_\theta; \theta \in \mathcal{O}\}$ was studied in Stute et al. (1993).
For one-dimensional discrete distributions, PB for testing goodness-of-fit was proposed by Henze (1996).

A natural question to ask is: Does parametric bootstrap work in a semiparametric setting?

The answer is yes, provided estimators are "regular" enough (Genest and Rémillard, 2008).

Basically, all known estimators are regular! Moreover, tests can be based on the empirical copula, Kendall's process, or Rosenblatt transform.

Example: Goodness-of-fit test for copulas

H_0: $C = C_{\theta_0}$, for some $\theta_0 \in \mathcal{O}$.

Parametric bootstrap works if (θ_n) is rank-based and regular.

Examples of regular estimators:

- If the pseudo Maximum Likelihood Estimator exists (Genest et al., 1995), it is regular.
- Other examples: Kendall's tau, Spearman's rho, and van der Waerden coefficient.

Parametric bootstrap implementation

Given (independent) random vectors $X_1, \ldots, X_n$, estimate θ by θ_n , compute the test statistic $S_n = \phi(\mathbb{C}_n)$, and repeat the following steps for $k = 1, \ldots, N$:

- Generate $U^*_{k,1}, \ldots, U^*_{k,n} \overset{i.i.d.}{\sim} C_{\theta_n}$.
- Estimate θ by $\theta^*_{k,n}$, using the sample $U^*_{k,1}, \ldots, U^*_{k,n}$.
- Compute $S^*_{k,n}\phi(\mathbb{C}^*_{k,n})$, where $\mathbb{C}^*_{k,n} = \sqrt{n}\left(C^*_{k,n} - C_{\theta^*_{k,n}}\right)$, and $C^*_{k,n}$ is the empirical copula of the sample $U^*_{k,1}, \ldots, U^*_{k,n}$.

The approximate P-value of S_n is computed as

$$\frac{1}{N}\sum_{k=1}^{N} \mathbf{1}(S^*_{k,n} > S_n),$$

provided large value of S_n indicates lack-of-fit.

Simulation experiment: GOF for the Gaussian copula

Table 2: Percentage of rejection of the null hypothesis of Gaussian copula for the Cramér-von Mises statistic S_n at the 5% level for Student copula alternatives, based on $N = 1000$ and $N = 10000$ replicates. The value of the unknown parameter ρ was 0.25.

Copula model	$n = 250$		$n = 500$	
	$N = 1000$	$N = 10000$	$N = 1000$	$N = 10000$
Gaussian	4.5	5.09	4.7	5.08
Student ($\nu = 20$)	7.4	6.45	6.7	6.13
Student ($\nu = 10$)	9.2	8.28	9.3	9.38
Student ($\nu = 5$)	14.8	14.88	19.8	21.20
Student ($\nu = 2.5$)	42.8	43.51	75.8	75.46
Student ($\nu = 2$)	63.3	63.17	94.1	94.22
Student ($\nu = 1.5$)	88.2	87.56	99.6	99.93

Insurance data example

Kendall's process $\mathbb{K}_n = n^{1/2}(K_n - K_{\theta_n})$ is used instead of the copula process, where $K(t) = P\{F(X_i) \le t\}$, while

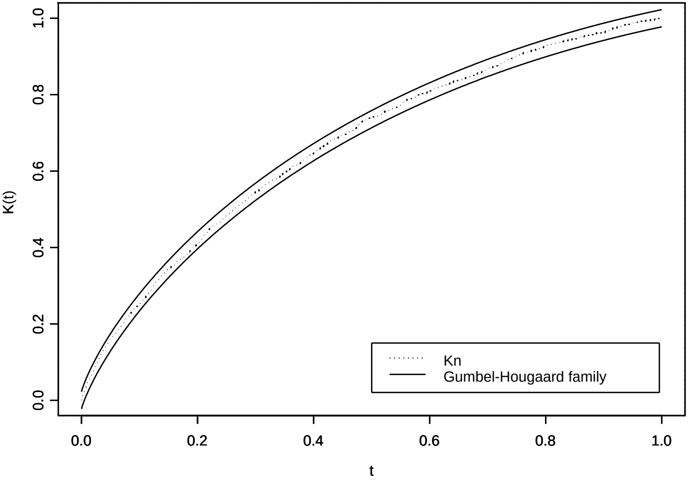
$$\hat{K}(t) = K_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{F_n(X_i) \le t\}.$$

Here, the test statistics are $S_n^{(K)} = \int_0^1 \mathbb{K}_n^2(t)dt$ and $T_n^{(K)} = \sup_{t\in[0,1]} |\mathbb{K}_n(t)|$.

Table 3: Results of the goodness-of-fit tests based on the statistics $S_n^{(K)}$, and $T_n^{(K)}$ for the data of LOSS and ALAE insurance data

Model	θ_n	$S_n^{(K)}$ / $T_n^{(K)}$	Critical value $c_{2n}(0.95)$	P-value (in %)
Clayton	0.939	2.284	0.129	0.0
		2.497	0.901	0.2
Frank	3.143	0.239	0.114	0.2
		0.893	0.856	3.1
Gumbel–Hougaard	0.319	0.028	0.117	86.5
		0.494	0.863	82.0

Figure 3: Estimator K_n of K for the LOSS and ALAE insurance data, with global 95% confidence bands based on $T_n^{(K)}$ for the Gumbel–Hougaard model.



Importance of the Rosenblatt transform

The Rosenblatt transform $\mathcal{R}$ of a copula C is the mapping from $(0,1)^d \mapsto (0,1)^d$, defined for $u = (u_1, \ldots, u_d)$, by

$$
\begin{aligned}
(\mathcal{R}(u))_1 &= u_1, \\
(\mathcal{R}(u))_j &= \frac{\frac{\partial^{j-1}}{\partial u_1 \cdots \partial u_{j-1}} C(u_1, \ldots, u_j, 1, \ldots, 1)}{\frac{\partial^{j-1}}{\partial u_1 \cdots \partial u_{j-1}} C(u_1, \ldots, u_{j-1}, 1, \ldots, 1)},
\end{aligned}
$$

$j \in \{2, \ldots, d\}$.

It is well-known that if $U \sim C$, then $\mathcal{R}(U) \sim C_\perp$, where $C_\perp$ is the independence copula.

It is often used for simulations. In fact, if $W \sim C_\perp$, i.e., W is uniformly distributed on $(0,1)^d$, then $U = \mathcal{R}^{-1}(W)$ has distribution C.

"Best" omnibus test for GOF of copulas

Instead of using $\mathbb{C}_n$ for testing $H_0 : C \in \{C_\theta; \theta \in \mathcal{O}\}$, define pseudo-observations $E_{1,n} = \mathcal{R}_{\theta_n}(e_{1,n})$, ..., $E_{n,n} = \mathcal{R}_{\theta_n}(e_{n,n})$, where R_θ is the Rosenblatt transform of C_θ.

Under the null hypothesis H_0, the empirical distribution function

$$D_n(u) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(E_{i,n} \le u), \quad u \in [0,1]^d,$$

associated with the pseudo-observations $E_1, \ldots, E_n$ should be "close" to the independence copula $C_\perp$.

According to Genest et al. (2009), the best omnibus test for goodness-of-fit is based on

$$S_n^{(B)} = n \int_{[0,1]^d} \{D_n(u) - C_\perp(u)\}^2 \, du.$$

P-values are computed using parametric bootstrap.

Parametric bootstrap and residuals of time series

According to Remillard (2010), tests of goodness-of-fit can be performed on the copula of innovations of time series models of the form $X_t = (X_{1t}, \ldots, X_{dt})$, where

$$X_{jt} = \mu_{jt}(\theta_0) + \sigma_{jt}(\theta_0)\epsilon_{jt}, \qquad j = 1, \ldots, d,$$

where the innovation process $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{dt})$ is a strong white noise and ϵ_t has copula C.

The technique used for serially independent data works without change!

Change-point tests are also valid for marginal distributions and copula of such models.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Example

Chen and Fan (2006) studied the dependence of the innovations for the Deutsche Mark/US and Japanese Yen/US exchanges rates, from April 28, 1988 to Dec 31, 1998.

AR(3)-GARCH(1,1) and AR(1)-GARCH(1,1) models were fitted on the 2684 log-returns.

Because the series are so long, univariate change-point tests were performed on the standardized residuals and the null hypothesis was accepted.

A copula change-point test was also performed, yielding a P-value of 33%.

Figure 2: Scatter plot of residuals.

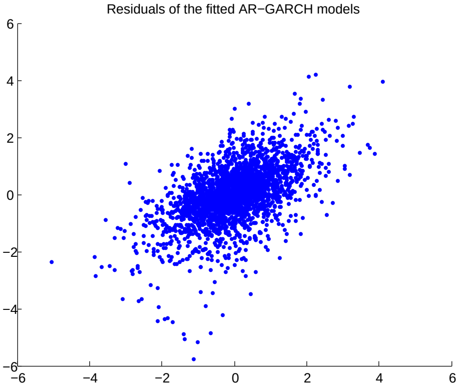
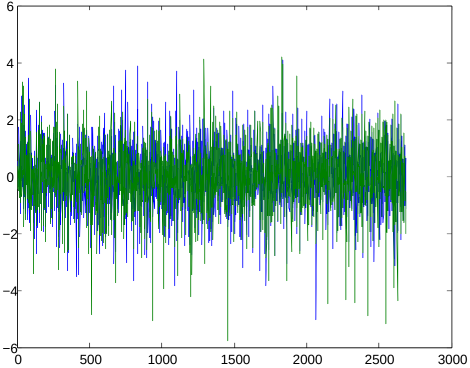


Figure 3: Residuals vs time.



Copula models considered by Chen and Fan (2006) (Gaussian, Student, Clayton, Frank, Gumbel) were all rejected using $S_n^{(B)}$, while they selected the Student copula as the best model, based on the likelihood rankings.

What is the model then?

The next best model would be a mixture of two Gaussian copulas (Dias and Embrechts, 2004).

H_0 was accepted with a 84% p-value, computed from $N = 100$ replications. The parameters of the two Gaussian copulas are $\hat{\rho} = [0.8205, 0.3749]$ and $\hat{\pi} = [0.4017, 0.5983]$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Parametric bootstrap for dynamic models

Given a time series Y_t, one may want to test

> *H_0: The conditional distribution of Y_t given $\mathcal{F}_{t-1}$ belongs to the parametric family $\{F_{t,\theta}; \theta \in \mathcal{O}\}$.*

One can show that PB can be used in p-Markov, ARMA, GARCH and regime-switching models.

One may also want to test semiparametric hypotheses like

> *H_0: The copula C of $(Y_{t-p}, \ldots, Y_t)$ belongs to the parametric family $\mathcal{C} = \{C_\phi; \phi \in \mathcal{P}\}$,*

It can be shown, for example that PB works for dynamic copulas, i.e., copulas associated with Markov processes.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Rosemblatt's transform is back!

Following an idea of Diebold et al. (1998),

> *H_0: For some $\theta \in \mathcal{O}$, the conditional distribution of Y_t given $\mathcal{F}_{t-1}$ is $F_{t,\theta}$, for all $t \geq 1$.*

is equivalent to

> *H_0: For some $\theta \in \mathcal{O}$, the Rosenblatt's transforms of Y_t given $\mathcal{F}_{t-1}$ is $R_{t,\theta}$, for all $t \geq 1$.*

Under H_0, $V_t = R_{t,\theta}(Y_t)$ are independent and uniformly distributed over $(0,1)^d$.

In practice, θ is unknown and estimated by θ_n, so V_t is replaced by $v_{n,t} = R_{t,\theta_n}(Y_t)$, $t \geq 1$.

Tests of goodness-of-fit could be based on $\mathbb{D}_n = \sqrt{n}(D_n - C_\perp)$, with $C_\perp(u) = \prod_{k=1} u_k$, $u = (u_1, \ldots, u_d) \in [0,1]^d$, and

$$D_n(u) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}(v_{n,t} \leq u) = \frac{1}{n}\sum_{t=1}^{n}\prod_{k=1}^{d} \mathbf{1}(v_{n,t,k} \leq u_k).$$

PB works if the estimator θ_n is "regular".

For example, MLE, moment and EM estimators are regular in general.

Gaussian regime-switching model for the S&P500 returns data

Here the null hypothesis is that conditionally on the (Markovian) regimes, the returns are independent and Gaussian.

Table 4: P-value for the goodness-of-fit test using 1000 replications.

Number of regimes	1	2	**3**	4
P-value	0	0	**9%**	3%

According to Table 4, one should choose a regime-switching model with 3 regimes, since it is the smallest number of regimes for which the P-value of the goodness-of-fit test is larger than 5%.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Dynamic copulas

Observation from a reader of the Globe and Mail:

> *The Canadian dollar always seems to go up [with respect to the US dollar] when oil prices rise. Is there a direct correlation between the two?*

With the exchange rate returns (X_t) and oil futures returns (Y_t), one would like to fit a 4-dimensional copula on $(X_t, X_{t-1}, Y_t, Y_{t-1})$.

By stationarity, the copulas associated with the pairs (X_t, Y_t) and (X_{t-1}, Y_{t-1}) are the same.

In practice, only the copula of the conditional distribution of (X_t, Y_t) given (X_{t-1}, Y_{t-1}) matters.

GOF tests for Oil vs Exchange rate

Clayton and Frank dynamic copula models were rejected, their p-values as approximated by PB being lower than 1%.

Table 5: Results of the estimation and goodness-of-fit for the dynamic Gaussian and Student copulas, using $N = 100$ iterations.

Period	*Gaussian*	*Student*
2008-2009	$\hat{\rho} = .435$, *PV = 12%*	$\hat{\rho} = .444$, $\hat{\nu} = 3.51$, *PV = 71%*
2005-2009	$\hat{\rho} = .350$, *PV = 53%*	$\hat{\rho} = .345$, $\hat{\nu} = 5.60$, *PV = 78%*
2000-2009	$\hat{\rho} = .236$, *PV = 1%*	$\hat{\rho} = .220$, $\hat{\nu} = 16.7$, *PV = 58%*

Table 6: Estimation and goodness-of-fit for the Student copulas, for the non-overlapping periods, using $N = 100$ iterations.

Period	*Student*
2008-2009	$\hat{\rho} = .444$, $\hat{\nu} = 3.51$, *PV = 71%*
2005-2007	$\hat{\rho} = .228$, $\hat{\nu} = 39.60$, *PV = 31%*
2000-2004	$\hat{\rho} = .086$, $\hat{\nu} = \infty$, *PV = 37%*

Conclusion I

- Parametric bootstrap is a powerful method that works fine but can be quite slow.
- The method is limited by available algorithms for calculating the DF under H_0.
- One could use Monte Carlo implementation! Replace the DF by an empirical one, obtained from independent Monte Carlo sampling. It is called two-level parametric bootstrap and it works but it is very very slow!
- No such computational problems for tests based on the Rosenblatt transform which is in addition almost always the best!
- Parametric bootstrap coupled with the Rosenblatt transform also works for dynamic models.

Multiplier CLT

An alternative Monte Carlo method, based on a version of the multiplier central limit theorem (van der Vaart and Wellner, 1996) has been proposed in some particular cases: Lin et al. (1993, 2002), Hansen (1996), Guay and Scaillet (2003), Scaillet (2005) and Rémillard and Scaillet (2009).

It seems difficult to find out when the multiplier methodology was first used in statistical inference.

As shown in the next example, it is sometime impossible to use the parametric bootstrap technique.

Consider a stochastic volatility model of the form $X_t = (X_{1t}, \ldots, X_{dt})$ such that

$$X_{jt} = \mu_{jt}(\theta_0) + \sigma_{jt}(\theta_0)\epsilon_{jt}, \qquad j = 1, \ldots, d,$$

where the innovation process $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{dt})$ is a strong white noise and ϵ_t has copula C.

In practice, the innovations are replaced by the residuals $e_{t,n}$.

Also, if one is interested in the dependence between the innovations, i.e., the copula C_t of ϵ_t, the pseudo-observations $U_{t,n} = (\mathrm{Rank}(e_{1t,n})/(n+1), \ldots, \mathrm{Rank}(e_{dt,n})/(n+1))$ are used to get rid of the unknown (continuous) marginal distribution od ϵ_{jt}.

According to Remillard (2010), a test of change-point in the copula,, i.e., $H_0 : C_1 = \cdots = C_n$ vs $H_1 : C_1 = \cdots C_{\tau-1} \neq C_\tau = \cdots = C_n$, for some τ, can be performed using the process

$$\mathbb{G}_n(s, u) = \frac{1}{\sqrt{n}} \sum_{t=1}^{\lfloor ns \rfloor} \left\{ \mathbf{1}(U_{t,n} \le u) - C_n(u) \right\},$$

which converges in law to

$$\mathbb{G}(s, u) = \mathbb{C}(s, u) - s\mathbb{C}(1, u) = \alpha(s, u) - s\alpha(1, u),$$

where α is a C-Kiefer process, whose law depends on C, which is unknown.

Recall that a C-Kiefer process is a continuous centered Gaussian process with covariance

$$\mathrm{Cov}\left\{\alpha(s, u), \alpha(t, v)\right\} = s \wedge t \left\{ C(u \wedge v) - C(u)C(v) \right\},$$

$s, t \in [0, 1]$, $u, v \in [0, 1]^d$.

The problem is then to generate independent copies of α.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

What is the multiplier CLT?

To illustrate the idea, take a random sample $Y_1, \ldots, Y_n$ of i.i.d. random variables with mean μ and variance σ^2. Then $Z_n = \sqrt{n}(\bar{Y} - \mu) \rightsquigarrow Z \sim N(0, \sigma^2)$.

Further let $\xi_1, \ldots, \xi_n \overset{i.i.d.}{\sim} N(0,1)$, also independent of $Y_1, \ldots, Y_n$. Set

$$\tilde{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i (Y_i - \bar{Y}).$$

Then $(Z_n, \tilde{Z}_n) \rightsquigarrow (Z, \tilde{Z})$, where $\tilde{Z} \sim N(0, \sigma^2)$ and $\tilde{Z}$ is independent of Z, i.e., $\tilde{Z}$ is an independent copy of Z. This is grosso modo the idea behind the multiplier central limit theorem.

Relation with traditional bootstrap

Traditional bootstrap is almost a special case of the multiplier technique.

Denoting by $\hat{Y}_1, \ldots, \hat{Y}_n$ the bootstrap sample, then $\overline{\hat{Y}} = \frac{1}{n}\sum_{i=1}^{n} N_i Y_i$, where N_i denotes the number of times Y_i appears in the bootstrap sample.

Therefore

$$\hat{Z}_n = \sqrt{n}\left(\overline{\hat{Y}} - \bar{Y}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(N_i - 1)(Y_i - \bar{Y})$$

and $\hat{Z}_n$ converges in law to an independent copy of Z.

Multiplier CLT for random vectors

The first result gives conditions for the “reproduction” of random vectors arising from non-observable functions.

Theorem 1: Suppose that $L : \mathfrak{X} \mapsto \mathbb{R}^d$ is a measurable function such that $\mathrm{E}\|L(X)\|^2 < \infty$ and $\mathrm{E}L(X) = 0$, and assume that

$$\Theta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} L(X_i) \rightsquigarrow \Theta.$$

Let L_n be an estimation of L and set

$$\tilde{\Theta}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i L_n(X_i),$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. with mean zero and variance one.

If $\|L_n - L\|_r \xrightarrow{Pr} 0$, then $(\Theta_n, \tilde{\Theta}_n) \rightsquigarrow (\Theta, \tilde{\Theta})$, where $\tilde{\Theta} \sim N\{0, \mathrm{cov}\ L(X)\}$, and $\tilde{\Theta}$ is independent of Θ.

In particular, if $\Theta \sim N\{0, \mathrm{cov}\ L(X)\}$, then $\tilde{\Theta}$ is an independent copy of Θ.

Implementation

In many statistical applications, one starts with an estimator

$$\theta_n = \frac{1}{n}\sum_{i=1} \ell_n(X_i)$$

of θ, and one wants to generate random values of Θ, where $\Theta_n = \sqrt{n}(\theta_n - \theta) \rightsquigarrow \Theta$.

In general, one cannot apply directly Theorem 1 to $\Theta_n = \sqrt{n}(\theta_n - \theta)$ by choosing $L_n = \ell_n$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Implementation steps

Before using Theorem 1, usually one has to work a little bit harder, by going through the following steps:

- Find L so that

$$\Theta_n = \frac{1}{\sqrt{n}} \sum_{i=1} L(X_i) + o_P(1)$$

and $\Theta_n \rightsquigarrow \Theta \sim N\{0, \mathrm{cov}\ L(X)\}$.

- Find L_n and $r \in \mathbb{R}_2^{\otimes d}$ such that $\|L_n - L\|_r \xrightarrow{Pr} 0$.

Outline

Motivation
Traditional bootstrap
Parametric bootstrap
Multipliers
References

Example

Suppose that $X_1 = (X_{11}, X_{12}), \ldots, X_n = (X_{n1}, X_{n2})$ are independent observations on a random vector X with distribution function H and margins F, G.

The empirical Kendall's coefficient can be expressed as

$$\tau_n = -1 + \frac{4}{n} \sum_{i=1}^{n} H_n(X_{i1}, X_{i2}) = 4\mu_n - 1,$$

where

$$H_n(x_1, x_2) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}(X_{j1} \leq x_1, X_{j2} \leq x_2)$$

is the empirical distribution function.

Step 1: From Ghoudi and Rémillard (2004),

$$\Theta_n = \sqrt{n}(\tau_n - \tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} L(X_i) + o_P(1),$$

with

$$L(x_1, x_2) = 8\left\{H(x_1, x_2) - \mu\right\} - 4F(x_1) - 4G(x_2) + 4,$$

$\mu = \mathrm{E}H(X)$, $\tau = 4\mu - 1$, $F(x_1) = H(x_1, \infty)$ and $G(x_2) = H(\infty, x_2)$.

Step 2: Take

$$L_n(x_1, x_2) = 8\{H_n(x_1, x_2) - \mu_n\} - 4F_n(x_1) - 4G_n(x_2) + 4,$$

with $F_n(x_1) = H_n(x_1, \infty)$ and $G_n(x_2) = H_n(\infty, x_2)$.

It then follows that

$$\sup_{x \in \mathbb{R}^2} |L_n(x) - L(x)| \leq 16 \sup_{x \in \mathbb{R}^2} |H_n(x) - H(x)| + 2|\tau_n - \tau|$$

converges to zero a.s., since $\{\mathbf{1}(-\infty, x]\,;\, x \in \mathbb{R}^2\}$ is a Glivenko-Cantelli class. Therefore one can take the supnorm.

Theorem 1 yields $(\Theta_n, \tilde{\Theta}_n) \rightsquigarrow (\Theta, \tilde{\Theta})$, where $\tilde{\Theta}$ is an independent copy of Θ.

Multiplier CLT for empirical processes

$e_{1,n}, \ldots, e_{n,n}$: random vectors defined from $X_1, \ldots, X_n$. For $t \in [-\infty, \infty]^d$, set

$$
\begin{aligned}
K_n(t) &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(e_{i,n} \leq t), \\
\tilde{\alpha}_n(t) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \{\mathbf{1}(e_{i,n} \leq t) - K_n(t)\}.
\end{aligned}
$$

Recall that a process $\mathbb{B}$ is a K-Brownian bridge if it is a continuous centered Gaussian process with covariance function $\mathrm{cov}\,\{\mathbb{B}(s), \mathbb{B}(t)\} = \min\{K(s), K(t)\} - K(s)K(t)$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Theorem 2: Suppose that K is a continuous DF on $\mathbb{R}^d$ and $K_n(t) \xrightarrow{Pr} K(t)$ for any fixed $t \in [-\infty, \infty]^d$.

Then $\tilde{\alpha}_n \rightsquigarrow \tilde{\alpha}$, where $\tilde{\alpha}$ is a K-Brownian bridge.

If in addition $\mathbb{K}_n = \sqrt{n}(K_n - K) \rightsquigarrow \mathbb{K}$, where $\mathbb{K}$ is a continuous centered Gaussian process, then $(\mathbb{K}_n, \tilde{\alpha}_n) \rightsquigarrow (\mathbb{K}, \tilde{\alpha})$ and $\tilde{\alpha}$ is independent of $\mathbb{K}$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Implementation

In most applications, $\mathbb{K}_n \rightsquigarrow \mathbb{K}$, and $\mathbb{K}$ is not a K-Brownian bridge. Instead, one can sometimes show (Ghoudi and Rémillard, 2004) that $\mathbb{K} = \alpha - \gamma$ where α is a K-Brownian bridge.

Therefore, using Theorem 2, one gets independent copies of α, not $\mathbb{K}$!

It remains to approximate the law of γ using Theorems 1 and 2.

Multipliers and GOF

As seen before, in many goodness-of-fit tests, the limiting empirical process is of the form $\mathbb{F} = \mathbb{B}_F - \Theta^\top \dot{F}$, where $\mathbb{B}_F$ is a F-Brownian bridge.

Then one can use multipliers to generate asymptotically independent copies of $(\alpha, \Theta, \mathbb{K})$.

Multipliers and copulas

Let $(X_i)_{i\geq 1}$ be i.i.d random vectors with continuous margins. Under regularity assumptions, $\mathbb{C}_n = \sqrt{n}(C_n - C) \rightsquigarrow \mathbb{C}$ where $\mathbb{C}(t) = \alpha(t) - \sum_{j=1}^{d} \partial_{t_j} C(t)\beta_j(t_j)$, with

$$\alpha_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [\mathbf{1}\{\epsilon_i \leq t\} - C(t)] \rightsquigarrow \alpha(t),$$

and $\beta_j(t_j) = \alpha(1, \ldots, 1, t_j, 1, \ldots, 1)$, for every $j \in \{1, \ldots, d\}$.

Here α is a C-Brownian bridge, and each β_j is a standard Brownian bridge.

For any $j \in \{1, \ldots, d\}$, set

$$\hat{f}_{jn}(t) = \widehat{\partial_{t_j} C(t)} = \frac{C_n(t + u_j/\sqrt{n}) - C_n(t - u_j/\sqrt{n})}{2/\sqrt{n}},$$

where $(u_j)_k = 1$ if $k = j$ and $(u_j)_k = 0$ otherwise.

It can be shown, e.g., Rémillard and Scaillet (2009), that for each $j \in \{1, \ldots, d\}$, $\hat{f}_{jn}(t)$ provides a (uniform) consistent estimate of $\partial_{t_j} C(t)$.

Therefore, if

$$\tilde{\alpha}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \left[\mathbf{1}\{e_{i,n} \le t\} - C_n(t) \right],$$

$$\tilde{\beta}_{jn}(t_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - \bar{\xi}) \mathbf{1}(e_{ji,n} \le t_j),$$

and

$$\tilde{C}_n(t) = \tilde{\alpha}_n(t) - \sum_{j=1}^{d} \hat{f}_{jn}(t) \tilde{\beta}_{jn}(t_j),$$

it follows from Theorem 2 that $(\mathbb{C}_n, \tilde{\mathbb{C}}_n) \rightsquigarrow (\mathbb{C}, \tilde{\mathbb{C}})$, where $\tilde{\mathbb{C}}$ is an independent copy of $\mathbb{C}$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

That result was used in Scaillet (2005) for tests of positive quadrant dependence.

An easy extension to (independent or paired) samples was suggested in Rémillard and Scaillet (2009) for testing equality between copulas.

Solution for the change-point test

To obtain asymptotically independent copies of the limiting test statistic $T_n = \max_{1\le j\le n} \max_{1\le i\le n} |\mathbb{G}_n(j/n, U_{i,n})|$, repeat the following steps for every $k \in \{1,\ldots,N\}$:

1. Generate a random sample $\xi_{i,k} \sim N(0,1)$, $i = 1,\ldots,n$.
2. For $(s,u) \in [0,1]^{d+1}$, set

$$\alpha_n^{(k)}(s,u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_{i,k}\{\mathbf{1}(U_{i,n} \le u) - C_n(u)\}$$

   and compute $\mathbb{G}_n^{(k)}(s,u) = \alpha_n^{(k)}(s,u) - \frac{\lfloor ns \rfloor}{n}\alpha_n^{(k)}(1,u)$.
3. Evaluate $T_n^{(k)} = \max_{1\le j\le n} \max_{1\le i\le n} \left|\mathbb{G}_n^{(k)}(j/n, U_{i,n})\right|$.

An approximate P-value for the test is then given by $\sum_{k=1}^{N} \mathbf{1}\left(S_n^{(k)} > S_n\right)/N$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Back to exchange rates data of Chen & Fan (2006)

A copula change-point test was performed using $N = 100$ replications. The estimated P-value was 33%.

It required 30 hours of calculations, using the multipliers methodology described before.

Multipliers and dynamic models

One has to be careful! Here is a an example, using a simple AR(1) model: $Y_t - \mu = \phi(Y_{t-1} - \mu) + \varepsilon_t$.

Suppose one wants to tests randomness in the AR(1) model, i.e, if the p consecutive innovations are independent, using residuals.

Let K be the DF of $\epsilon = (\varepsilon_1, \ldots, \varepsilon_p)$, where $p > 1$. Under the H_0, one can estimate $K(t) = F(t_1) \cdots F(t_p)$ by the empirical distribution function K_n of the residuals, i.e.

$$K_n(t) = \frac{1}{n} \sum_{i=1}^{n-p+1} \mathbf{1}(e_{i,n} \leq t_1, \ldots, e_{i+p-1,n} \leq t_p).$$

Hence, tests of randomness could be based on the empirical process

$$\sqrt{n} \left\{ K_n(t_1, \ldots, t_p) - \prod_{k=1}^{p} F_n(t_k) \right\}.$$

See, e.g., Ghoudi et al. (2001).

Outline

Motivation
Traditional bootstrap
Parametric bootstrap
Multipliers
References

Under extra conditions, $\mathbb{K}_n = \sqrt{n}\{K_n - K\} \rightsquigarrow \mathbb{K}$, where

$$\begin{aligned}\mathbb{K}(t) &= \alpha(t) + (1-\phi)\mathbb{M}\sum_{j=1}^{p} F'(t_j)\prod_{l\neq j} F(t_l) \\ &\quad + \Phi\sum_{j=1}^{p} F'(t_j)\sum_{q=j+1}^{p}\phi^{q-j-1}G(t_q)\prod_{i\neq j,q} F(t_i),\end{aligned}$$

with $G(s) = \mathrm{E}\{\varepsilon_1 \mathbf{1}(\varepsilon_1 \leq s)\}$, and $\alpha_n \rightsquigarrow \alpha$, where

$$\alpha_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n-p+1}\{\mathbf{1}(\varepsilon_i \leq t_1, \ldots, \varepsilon_{i+p-1} \leq t_p) - K(t)\}.$$

Note that α is not a K-Brownian bridge.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Taking care of the parameters

To use Theorem 1, set $X_i = (Y_{i-1}, Y_i)$,

$$
\begin{aligned}
H(x_1, x_2) &= x_2 - \mu - \phi(x_1 - \mu), \\
H_n(x_1, x_2) &= x_2 - \mu_n - \phi_n(x_1 - \mu_n), \\
\mathbb{M}_n &= n^{1/2}(\mu_n - \mu) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{H(X_i)}{1-\phi} + o_P(1), \\
\Phi_n &= n^{1/2}(\phi_n - \phi) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} L(X_i) + o_P(1),
\end{aligned}
$$

where $L(x_1, x_2) = H(x_1, x_2)(x_1 - \mu)$, and $L_n(x_1, x_2) = H_n(x_1, x_2)(x_1 - \mu_n)$.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Taking care of α

Set $p = 2$ for simplicity. Then

$$\alpha_n(t_1, t_2) \;=\; \gamma_n(t_1, t_2) + F(t_1)\beta_n(t_2) + F(t_2)\beta_n(t_1) + o_P(1),$$

where

$$\beta_n(t_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\mathbf{1}(\varepsilon_i \le t_1) - F(t_1)\},$$

and

$$\gamma_n(t_1, t_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\mathbf{1}(\varepsilon_i \le t_1) - F(t_1)\}\{\mathbf{1}(\varepsilon_{i+1} \le t_2) - F(t_2)\}.$$

It follow from Theorem 2 that if one sets

$$\tilde{\gamma}_n(t_1, t_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \prod_{j=1}^{2} \{\mathbf{1}(e_{i+j-1,n} \le t_j) - F_n(t_j)\}$$

$$\tilde{\beta}_n(t_1, t_2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i \{\mathbf{1}(e_{i,n} \le t_1) - F_n(t_1)\},$$

then

$$\tilde{\alpha}_n(t_1, t_2) = \tilde{\gamma}_n(t_1, t_2) + F_n(t_1)\tilde{\beta}_n(t_2) + F_n(t_2)\tilde{\beta}_n(t_1)$$

will converge in law to an independent copy $\tilde{\alpha}$ of α.
See Rémillard (2011) for more details.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

Conclusion II

- The multiplier method works in more general settings than the parametric bootstrap.
- It is faster than traditional and parametric bootstrap methods but requires more work.
- Can be also used with residuals of dynamic models for testing change-point, randomness and goodness-of-fit.
- Comparisons in terms of power are yet to be made.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

References I

Berrada, T., Dupuis, D. J., Jacquier, E., Papageorgiou, N., and Rémillard, B. (2006). Credit migration and derivatives pricing using copulas. *J. Comput. Fin.*, 10:43–68.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics*, 135(1-2):125 – 154.

Dias, A. and Embrechts, P. (2004). Dynamic copula models for multivariate high-frequency data in finance. Technical report, ETH Zürich.

Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, 39(4):863–883.

References II

Fermanian, J.-D., Radulović, D., and Wegkamp, M. J. (2004). Weak convergence of empirical copula processes. *Bernoulli*, 10:847–860.

Genest, C., Ghoudi, K., and Rivest, L.-P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 82:543–552.

Genest, C., Quessy, J.-F., and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the integral probability transformation. *Scand. J. Statist.*, 33:337–366.

Genest, C. and Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. *Ann. Inst. H. Poincaré Sect. B*, 44:1096–1127.

References III

Genest, C., Rémillard, B., and Beaudoin, D. (2009). Omnibus goodness-of-fit tests for copulas: A review and a power study. *Insurance Math. Econom.*, 44:199–213.

Ghoudi, K., Kulperger, R. J., and Rémillard, B. (2001). A nonparametric test of serial independence for time series and residuals. *J. Multivariate Anal.*, 79:191–218.

Ghoudi, K. and Rémillard, B. (2004). Empirical processes based on pseudo-observations. II. The multivariate case. In *Asymptotic Methods in Stochastics*, volume 44 of *Fields Inst. Commun.*, pages 381–406. Amer. Math. Soc., Providence, RI.

Guay, A. and Scaillet, O. (2003). Inderect inference, nuisance parameter, and threshold moving average models. *Journal of Business and Economic Statistics*, 21(1):122–132.

References IV

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 73:413–430.

Henze, N. (1996). Empirical-distribution-function goodness-of-fit tests for discrete models. *Canad. J. Statist.*, 24:81–93.

Lin, D. Y., Wei, L. J., and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572.

Lin, D. Y., Wei, L. J., and Ying, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics*, 58(1):1–12.

Papageorgiou, N., Rémillard, B., and Hocquard, A. (2008). Replicating the properties of hedge fund returns. *Journal of Alternative Invesments*, 11:8–38.

References V

Remillard, B. (2010). Goodness-of-fit tests for copulas of multivariate time series. Working paper series.

Rémillard, B. (2011). The multiplier central limit theorem for pseudo-observations and its applications. Technical Report G-2011-??, GERAD.

Remillard, B., Hocquard, A., and Papageorgiou, N. A. (2010). Option Pricing and Dynamic Discrete Time Hedging for Regime-Switching Geometric Random Walks Models. Working paper series.

Rémillard, B. and Scaillet, O. (2009). Testing for equality between two copulas. *J. Multivariate Anal.*, 100:377–386.

Scaillet, O. (2005). A Kolmogorov-Smirnov type test for positive quadrant dependence. *Canad. J. Statist.*, 33(3):415–427.

Outline

Motivation

Traditional bootstrap

Parametric bootstrap

Multipliers

References

References VI

Stute, W., Gonzáles Manteiga, W., and Presedo Quindimil, M. (1993). Bootstrap based goodness-of-fit tests. *Metrika*, 40:243–256.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York.