ENTROPY-BASED TEST FOR TIME SERIES MODELS

Siyun Park

Korea University Business School, Seoul, Korea

Sangyeol Lee and Jiyeon Lee

Department of Statistics, Seoul National University

Outline

- Introduce entropy-based test of fit in iid sample
  - Test statistic and the asymptotic distribution
  - Practical issues
- Extend and apply the test to time series models
  - Autoregressive models
  - GARCH models
  - Simulation result and application to real data

Maximum entropy principle

- Shannon entropy :

  the average unpredictability in a random variable,

  its information content

- The maximum entropy principle(Janes, 1957):

  Its applications successfully proven in various fields,

  computer vision, natural language processing.

Entropy

Boltzmann-Shannon entropy:

$$H(f) = -\int_{-\infty}^{\infty} f(x)\log(f(x))dx. \qquad (1)$$

Forte and Hughes (1988) proposed the function

$$\bar{H} = -\sum p_i \log(p_i/(x_i - x_{i-1})) \qquad (2)$$

as the discrete analogue of (1), $p_i = \int_{x_{i-1}}^{x_i} f(x)dx$.

for a variable defined in $[a, b]$,

$$\lim_{\max_i |x_i - x_{i-1}| \to 0} - \sum_{i=1}^{n} p_i \log(p_i/(x_i - x_{i-1})) = H(f), \tag{3}$$

where $p_i = P[x_{i-1} < X \leq x_i] = \int_{x_{i-1}}^{x_i} f(x)dx$, $i = 1, \ldots, n-1$ and $a = x_0 < \ldots < x_n = b$.

Simple null hypothesis case

Let Y_i, $i = 1, \ldots, n$ be a random sample from F,

$$H_0 : F = F_0 \quad \text{vs.} \quad F \neq F_0.$$

Generalized entropy: Lee et al.(2011)

A generalization of Forte and Hughes (1988)'s:

$$S^w(F) = -\sum_{i=1}^{m} w_i \left(F(s_i) - F(s_{i-1})\right) \log\left(\frac{F(s_i) - F(s_{i-1})}{s_i - s_{i-1}}\right), \qquad (4)$$

where the w's are appropriate weight functions with $0 \leq w_i \leq 1$ and $\sum_{i=1}^{m} w_i = 1$, m is the number of disjoint intervals for partitioning the data range, and $-\infty < a \leq s_1 \leq \ldots \leq s_m \leq b < \infty$ are preassigned partition points.

Test statistic

the null hypothesis will be rejected if

$$|S^w(F_n) - S^w(F_0)| \geq c$$

or, even more stringently, if

$$\sup_w |S^w(F_n) - S^w(F_0)| \geq c,$$

where

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I(Y_i \leq x).$$

Probability integral transformation

$$U_i = F_0(Y_i),$$

$$H_0 : F = F_0 \equiv U[0,1] \quad \text{vs.} \quad H_1 : F \neq F_0 \equiv U[0,1].$$

If F_0 is uniform distribution then $S^w(F_0) = 0$.

Use U_i and $F_n(u) = n^{-1} \sum_{i=1}^{n} I(U_i \leq u)$.

Probability integral transformation

$$U_i = F_0(Y_i),$$

$$H_0 : F = F_0 \equiv U[0,1] \quad \text{vs.} \quad H_1 : F \neq F_0 \equiv U[0,1].$$

If F_0 is uniform distribution then $S^w(F_0) = 0$.

Use U_i and $F_n(u) = n^{-1} \sum_{i=1}^{n} I(U_i \leq u)$.

Theorem (Lee et al.(2011))

Under H_0, as $n \to \infty$,

$$\sqrt{n} \sup_{\{w \in W\}} |S^w(F_n)| \xrightarrow{d} \sup_{\{w \in W\}} \Big| \sum_{i=1}^{m} w_i \left(BB(s_i) - BB(s_{i-1})\right) \Big|,$$

where $BB(s)$ is the Brownian bridge on [0,1], W denotes the space of bounded weight functions $w_i : [0,1] \to [0,1]$ with $\sum_{i=1}^{m} w_i = 1$, and $0 = s_0 \leq s_1 \leq \ldots \leq s_m = 1$.

Composite null hypothesis case

Y_i, $i = 1, \ldots, n$ be a random sample from F

$$H_0 : F \in \{F_0(x;\theta); \theta \in \Theta\} \quad \text{vs.} \quad H_1 : \text{not } H_0,$$

F_0: continuous distribution,

Θ: a d-dimensional parameter space.

- Let $\hat{\theta}_n$ with $n^{1/2}(\hat{\theta}_n - \theta_0) = O_P(1)$ under H_0.

- Let $U_i = F_0(Y_i; \theta_0)$, $\hat{U}_i = F_0(Y_i; \hat{\theta}_n)$,

$$F_n(s) = \frac{1}{n}\sum_{i=1}^{n} I(U_i \leq s) \text{ and } \hat{F}_n(s) = \frac{1}{n}\sum_{i=1}^{n} I(\hat{U}_i \leq s), \ s \in [0,1].$$

- Define

  the empirical process: $\mathcal{E}_n(s) = n^{1/2}(F_n(s) - s)$

  the estimated empirical process: $\hat{\mathcal{E}}_n(s) = n^{1/2}(\hat{F}_n(s) - s)$.

- We can express

$$\hat{\mathcal{E}}_n(s) = \mathcal{E}_n(s) - h(s)' \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1),$$

where $h(s) = \frac{\partial F_{\theta_0}(F_{\theta_0}^{-1}(s))}{\partial \theta}$.

- It can be seen that

$$\begin{aligned} |\sqrt{n} S^w(\hat{F}_n)| &= \Bigg| \sum_{i=1}^m w_i[\mathcal{E}_n(s_i) - \mathcal{E}_n(s_{i-1})] \\ &+ \sqrt{n}(\hat{\theta}_n - \theta_0)' \sum_{i=1}^m w_i[h(s_{i-1}) - h(s_i)] \Bigg| + o_P(1). \end{aligned}$$

- Hence, the previous theorem can be applied when $\max_{1 \le i \le m} |s_i - s_{i-1}| \to 0$ and $m \to \infty$.

- We can express

$$\hat{\mathcal{E}}_n(s) = \mathcal{E}_n(s) - h(s)' \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1),$$

where $h(s) = \frac{\partial F_{\theta_0}(F_{\theta_0}^{-1}(s))}{\partial \theta}$.

- It can be seen that

$$\begin{aligned} |\sqrt{n} S^w(\hat{F}_n)| &= \Bigg| \sum_{i=1}^{m} w_i [\mathcal{E}_n(s_i) - \mathcal{E}_n(s_{i-1})] \\ &+ \sqrt{n}(\hat{\theta}_n - \theta_0)' \sum_{i=1}^{m} w_i [h(s_{i-1}) - h(s_i)] \Bigg| + o_P(1). \end{aligned}$$

- Hence, the previous theorem can be applied when $\max_{1 \le i \le m} |s_i - s_{i-1}| \to 0$ and $m \to \infty$.

- We can express

$$\hat{\mathcal{E}}_n(s) = \mathcal{E}_n(s) - h(s)' \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1),$$

where $h(s) = \frac{\partial F_{\theta_0}(F_{\theta_0}^{-1}(s))}{\partial \theta}$.

- It can be seen that

$$\begin{aligned} |\sqrt{n} S^w(\hat{F}_n)| &= \left| \sum_{i=1}^{m} w_i [\mathcal{E}_n(s_i) - \mathcal{E}_n(s_{i-1})] \right. \\ &+ \left. \sqrt{n}(\hat{\theta}_n - \theta_0)' \sum_{i=1}^{m} w_i [h(s_{i-1}) - h(s_i)] \right| + o_P(1). \end{aligned}$$

- Hence, the previous theorem can be applied when $\max_{1 \le i \le m} |s_i - s_{i-1}| \to 0$ and $m \to \infty$.

Theorem

Under H_0*, as* $n \to \infty$*, if* $\max_{1 \le i \le m} |s_i - s_{i-1}| \to 0$ *and* $m \to \infty$,

$$\sqrt{n} \sup_{\{w \in W\}} |S^w(\hat{F}_n)| \xrightarrow{d} \sup_{\{w \in W\}} \Big| \sum_{i=1}^{m} w_i \left(BB(s_i) - BB(s_{i-1}) \right) \Big|,$$

where $BB(s)$ *is the Brownian bridge on [0,1],* W *denotes the space of bounded weight functions* $w_i : [0,1] \to [0,1]$ *with* $\sum_{i=1}^{m} w_i = 1$*, and* $0 = s_0 \le s_1 \le \ldots \le s_m = 1$.

Practical issues: $\sup_w$

Generate $w_i^{(l)}, l = 1, \ldots, L,$ iid $\sim U[0, 1]$. Put

$$w_{li} = \frac{w_i^{(l)}}{w_1^{(l)} + \cdots + w_m^{(l)}}$$

-Lee et al.(2011) : as $L \to \infty$,

$$\max_{1 \le l \le L} \Big| \sum_{i=1}^{m} w_{li} \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \Big| \xrightarrow{d} \sup_{w \in W} \left| \sum_{i=1}^{m} w_i \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right| .$$

- L=1000 is recommended.

Practical issues: $\sup_w$

Generate $w_i^{(l)}, l = 1, \ldots, L,$ iid $\sim U[0,1]$. Put

$$w_{li} = \frac{w_i^{(l)}}{w_1^{(l)} + \cdots + w_m^{(l)}}$$

-Lee et al.(2011) : as $L \to \infty$,

$$\max_{1 \le l \le L} \left| \sum_{i=1}^{m} w_{li} \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right| \xrightarrow{d} \sup_{w \in W} \left| \sum_{i=1}^{m} w_i \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right|.$$

- L=1000 is recommended.

Practical issues: $\sup_w$

Generate $w_i^{(l)}, l = 1, \ldots, L,$ iid $\sim U[0,1]$. Put

$$w_{li} = \frac{w_i^{(l)}}{w_1^{(l)} + \cdots + w_m^{(l)}}$$

-Lee et al.(2011) : as $L \to \infty$,

$$\max_{1 \le l \le L} \left| \sum_{i=1}^m w_{li} \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right| \xrightarrow{d} \sup_{w \in W} \left| \sum_{i=1}^m w_i \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right| .$$

- L=1000 is recommended.

Practical issues: $\sup_w$

Generate $w_i^{(l)}, l = 1, \ldots, L,$ iid $\sim U[0,1]$. Put

$$w_{li} = \frac{w_i^{(l)}}{w_1^{(l)} + \cdots + w_m^{(l)}}$$

-Lee et al.(2011) : as $L \to \infty$,

$$\max_{1 \le l \le L} \left| \sum_{i=1}^{m} w_{li} \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right| \xrightarrow{d} \sup_{w \in W} \left| \sum_{i=1}^{m} w_i \left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1}) \right) \right| .$$

- L=1000 is recommended.

Practical Issues: s_i, m

Take $s_i = i/m$, $i = 1, \ldots, m$.

Conventionally, m is chosen to be much less than n so that $m/n \to 0$. $m \approx n^{1/3}$ is recommended.

Test statistic:

$$\begin{aligned} \hat{T}_n &= \sqrt{n} \max_{1\le l\le L} \Big| \sum_{i=1}^{m} w_{li}(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \\ &\times \log m(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \Big|. \end{aligned}$$

Practical Issues: s_i, m

Take $s_i = i/m$, $i = 1, \ldots, m$.

Conventionally, m is chosen to be much less than n so that $m/n \to 0$. $m \approx n^{1/3}$ is recommended.

Test statistic:

$$\hat{T}_n = \sqrt{n} \max_{1 \le l \le L} \left| \sum_{i=1}^{m} w_{li} (\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \times \log m(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \right|.$$

Practical Issues: s_i, m

Take $s_i = i/m$, $i = 1, \ldots, m$.

Conventionally, m is chosen to be much less than n so that $m/n \to 0$. $\quad m \approx n^{1/3}$ is recommended.

Test statistic:

$$\hat{T}_n = \sqrt{n} \max_{1 \le l \le L} \Big| \sum_{i=1}^{m} w_{li} (\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \times \log m(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \Big|.$$

Practical Issues: s_i, m

Take $s_i = i/m$, $i = 1, \ldots, m$.

Conventionally, m is chosen to be much less than n so that $m/n \to 0$. $m \approx n^{1/3}$ is recommended.

Test statistic:

$$\hat{T}_n = \sqrt{n} \max_{1 \le l \le L} \Bigg| \sum_{i=1}^{m} w_{li} (\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \times \log m(\hat{F}_n(i/m) - \hat{F}_n((i-1)/m)) \Bigg| .$$

Wrap up and Move on

- The entropy-based GOF test in iid sample

  - Valid to various distributions, a variety of values of m

  - Can be applied in a variety of fields

AR model

Autoregressive model:

$$X_t - \beta_1 X_{t-1} - \cdots - \beta_q X_{t-q} = \epsilon_t, \qquad (5)$$

where ϵ_t are iid r.v.s with $E\epsilon_1 = 0$, $E{\epsilon_1}^2 = \sigma^2$, and $E{\epsilon_1}^4 < \infty$.

$$\begin{aligned}\phi(z) &= 1 - \beta_1 z - \cdots - \beta_q z^q \\ &= (1-z)^a (1+z)^b \prod_{k=1}^{l} (1 - 2\cos\theta_k z + z^2)^{d_k} \psi(z),\end{aligned}$$

where a, b, l, d_k are nonnegative integers, θ_k belongs to $(0, \pi)$ and $\psi(z)$ is the polynomial of order $r = q - (a + b + 2d_1 + \ldots + 2d_l)$ that has no zeros on the unit disk in the complex plane.

- If a, b, l, d_k are all zeros, $\{X_t\}$ is a stationary process.

AR model

Autoregressive model:

$$X_t - \beta_1 X_{t-1} - \cdots - \beta_q X_{t-q} = \epsilon_t, \tag{5}$$

where ϵ_t are iid r.v.s with $E\epsilon_1 = 0$, $E{\epsilon_1}^2 = \sigma^2$, and $E{\epsilon_1}^4 < \infty$.

$$\begin{aligned}\phi(z) &= 1 - \beta_1 z - \cdots - \beta_q z^q \\ &= (1-z)^a (1+z)^b \prod_{k=1}^{l} (1 - 2\cos\theta_k z + z^2)^{d_k} \psi(z),\end{aligned}$$

where a, b, l, d_k are nonnegative integers, θ_k belongs to $(0, \pi)$ and $\psi(z)$ is the polynomial of order $r = q - (a + b + 2d_1 + \ldots + 2d_l)$ that has no zeros on the unit disk in the complex plane.

- If a, b, l, d_k are all zeros, $\{X_t\}$ is a stationary process.

AR model

Autoregressive model:

$$X_t - \beta_1 X_{t-1} - \cdots - \beta_q X_{t-q} = \epsilon_t, \tag{5}$$

where ϵ_t are iid r.v.s with $E\epsilon_1 = 0$, $E{\epsilon_1}^2 = \sigma^2$, and $E{\epsilon_1}^4 < \infty$.

$$\begin{aligned} \phi(z) &= 1 - \beta_1 z - \cdots - \beta_q z^q \\ &= (1-z)^a (1+z)^b \prod_{k=1}^{l} (1 - 2\cos\theta_k z + z^2)^{d_k} \psi(z), \end{aligned}$$

where a, b, l, d_k are nonnegative integers, θ_k belongs to $(0, \pi)$ and $\psi(z)$ is the polynomial of order $r = q - (a + b + 2d_1 + \ldots + 2d_l)$ that has no zeros on the unit disk in the complex plane.

- If a, b, l, d_k are all zeros, $\{X_t\}$ is a stationary process.

the null hypothesis

$$H_0 : \epsilon_t \sim F_0(\cdot/\sigma), \sigma > 0 \quad \text{vs.} \quad H_1 : \text{not } H_0$$

- LSE $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_q)'$

  - the LSE has a limiting distribution of a functional form of standard Brownian motions (cf. Chan and Wei, 1988).

- the residual empirical process:

$$\hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s) \quad \text{with} \quad \hat{F}_n(s) = \frac{1}{n}\sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t/\hat{\sigma}_n) \le s)$$

  where $\hat{\epsilon}_t = X_t - \hat{\beta}_n' \mathbf{X}_{t-1}$ and $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t{}^2$.

- LSE $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0 = (\beta_1, \ldots, \beta_q)'$

  - the LSE has a limiting distribution of a functional form of standard Brownian motions (cf. Chan and Wei, 1988).

- the residual empirical process:

$$\hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s) \quad \text{with} \quad \hat{F}_n(s) = \frac{1}{n}\sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t/\hat{\sigma}_n) \leq s)$$

where $\hat{\epsilon}_t = X_t - \hat{\boldsymbol{\beta}}_n' \mathbf{X}_{t-1}$ and $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^{n} \hat{\varepsilon}_t{}^2$.

Lee and Wei (1999)

$$\begin{aligned}\hat{\mathcal{V}}_n(s) &= \mathcal{V}_n(s) - (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbf{X}_{t-1} f_0(F_0^{-1}(s)) \\ &+ \frac{\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)}{2\sigma_0^2} f_0(F_0^{-1}(s)) F_0^{-1}(s) + \Delta_n(s),\end{aligned}$$

where σ_0 and $\boldsymbol{\beta}_0$ denote the true values under H_0,
$\mathcal{V}_n(s) = \sqrt{n}(F_n(s) - s)$, $F_n(s) = \frac{1}{n}\sum_{t=1}^{n} I(F_0(\epsilon_t/\sigma_0) \leq s)$, $f_0 = F_0'$,
and $\sup_s |\Delta_n(s)| = o_P(1)$.

Theorem

If f_0 *satisfies* $\lim_{|x|\to\infty} |xf_0(x)| = 0$ and $\sup_x |f_0^{'}(x)| < \infty$, *and if* $\max_{1\le i\le m} |s_i - s_{i-1}| \to 0$ *and* $m \to \infty$, *as* $n \to \infty$, *under* H_0,

$$\sqrt{n} \sup_{\{w\in W\}} |S^w(\hat{F}_n)| \overset{d}{\approx} \sup_{\{w\in W\}} \Big| \sum_{i=1}^{m} w_i \left(BB(s_i) - BB(s_{i-1})\right) \Big|,$$

where W *denotes the space of bounded weight functions* $w : [0,1] \to [0,1]$ *with* $\sum_{i=1}^{m} w_i = 1$, *and* $0 = s_0 \le s_1 \le \ldots \le s_m = 1$.

Bootstrap method: Stute et al.(1993)

Step 1: Obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ based on the data $X_1, \ldots, X_n$.

Step 2: Generate $\epsilon_1^*, \ldots, \epsilon_n^* \sim F_0(\cdot/\hat{\sigma}_n)$ and construct $X_1^*, \ldots X_n^*$ through the model with the LSE by letting $X_i^* = 0$ for all $i \leq 0$. Then, calculate the test statistic, $\hat{T}_n^*$ with m based on these r.v.s.

Step 3: Repeat the above procedure, B times and calculate the $100(1-\alpha)$% percentile of the obtained B number of $\hat{T}_n^*$ values.

Step 4: Reject H_0 if the $\hat{T}_n$ value based on the observations is larger than the obtained $100(1-\alpha)$% percentile in Step3.

Bootstrap method: Stute et al.(1993)

Step 1: Obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ based on the data $X_1, \ldots, X_n$.

Step 2: Generate $\epsilon_1^*, \ldots, \epsilon_n^* \sim F_0(\cdot/\hat{\sigma}_n)$ and construct $X_1^*, \ldots X_n^*$ through the model with the LSE by letting $X_i^* = 0$ for all $i \leq 0$. Then, calculate the test statistic, $\hat{T}_n^*$ with m based on these r.v.s.

Step 3: Repeat the above procedure, B times and calculate the $100(1-\alpha)\%$ percentile of the obtained B number of $\hat{T}_n^*$ values.

Step 4: Reject H_0 if the $\hat{T}_n$ value based on the observations is larger than the obtained $100(1-\alpha)\%$ percentile in Step3.

Bootstrap method: Stute et al.(1993)

Step 1: Obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ based on the data $X_1, \ldots, X_n$.

Step 2: Generate $\epsilon_1^*, \ldots, \epsilon_n^* \sim F_0(\cdot/\hat{\sigma}_n)$ and construct $X_1^*, \ldots X_n^*$ through the model with the LSE by letting $X_i^* = 0$ for all $i \leq 0$. Then, calculate the test statistic, $\hat{T}_n^*$ with m based on these r.v.s.

Step 3: Repeat the above procedure, B times and calculate the $100(1-\alpha)$% percentile of the obtained B number of $\hat{T}_n^*$ values.

Step 4: Reject H_0 if the $\hat{T}_n$ value based on the observations is larger than the obtained $100(1-\alpha)$% percentile in Step3.

Bootstrap method: Stute et al.(1993)

Step 1: Obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ based on the data $X_1, \ldots, X_n$.

Step 2: Generate $\epsilon_1^*, \ldots, \epsilon_n^* \sim F_0(\cdot/\hat{\sigma}_n)$ and construct $X_1^*, \ldots X_n^*$ through the model with the LSE by letting $X_i^* = 0$ for all $i \leq 0$. Then, calculate the test statistic, $\hat{T}_n^*$ with m based on these r.v.s.

Step 3: Repeat the above procedure, B times and calculate the 100$(1-\alpha)$% percentile of the obtained B number of $\hat{T}_n^*$ values.

Step 4: Reject H_0 if the $\hat{T}_n$ value based on the observations is larger than the obtained 100$(1-\alpha)$% percentile in Step3.

Bootstrap method: Stute et al.(1993)

Step 1: Obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ based on the data $X_1, \ldots, X_n$.

Step 2: Generate $\epsilon_1^*, \ldots, \epsilon_n^* \sim F_0(\cdot/\hat{\sigma}_n)$ and construct $X_1^*, \ldots X_n^*$ through the model with the LSE by letting $X_i^* = 0$ for all $i \leq 0$. Then, calculate the test statistic, $\hat{T}_n^*$ with m based on these r.v.s.

Step 3: Repeat the above procedure, B times and calculate the $100(1-\alpha)$% percentile of the obtained B number of $\hat{T}_n^*$ values.

Step 4: Reject H_0 if the $\hat{T}_n$ value based on the observations is larger than the obtained $100(1-\alpha)$% percentile in Step3.

Simulation study

Generate AR(1) processes,

$X_t = \beta X_{t-1} + \epsilon_t$, $\epsilon_t \sim$ iid $N(0, 1)$

$X_t = \beta X_{t-1} + \epsilon_t$, $\epsilon_t \sim$ iid $\epsilon_t \sim p\, N(0, 1) + (1 - p)\, N(0, \sigma_1^2)$

for $\beta = 0.1, 0.3, 0.5, 1.0$, $p = 0.9$, and $\sigma_1^2 = 10, 25$

Replicate 1000 times with $B = 500$ at the nominal level 0.05

Simulation study

Generate AR(1) processes,

$X_t = \beta X_{t-1} + \epsilon_t$, $\epsilon_t \sim$ iid $N(0, 1)$

$X_t = \beta X_{t-1} + \epsilon_t$, $\epsilon_t \sim$ iid $\epsilon_t \sim p\, N(0, 1) + (1 - p)\, N(0, \sigma_1^2)$

for $\beta = 0.1, 0.3, 0.5, 1.0$, $p = 0.9$, and $\sigma_1^2 = 10, 25$

Replicate 1000 times with $B = 500$ at the nominal level 0.05

the empirical sizes

β	$n=100$ $m=3$	$n=300$ $m=5$	$n=500$ $m=7$
0.3	0.049	0.056	0.053
0.5	0.046	0.043	0.057
0.7	0.043	0.051	0.051
1.0	0.051	0.048	0.044

- the test has no size distortions.

- also stable size for large $n(\geq 500)$.

the empirical sizes

β	$n = 100$ $m = 3$	$n = 300$ $m = 5$	$n = 500$ $m = 7$
0.3	0.049	0.056	0.053
0.5	0.046	0.043	0.057
0.7	0.043	0.051	0.051
1.0	0.051	0.048	0.044

- the test has no size distortions.

- also stable size for large $n(\geq 500)$.

the empirical powers

σ_1^2	β	$n = 100$ $m = 3$	$n = 300$ $m = 5$	$n = 500$ $m = 7$
10	0.3	0.386	0.771	0.942
	0.5	0.368	0.767	0.928
	0.7	0.394	0.747	0.933
	1.0	0.410	0.787	0.954
25	0.3	0.823	0.997	1.000
	0.5	0.828	0.998	1.000
	0.7	0.852	0.998	1.000
	1.0	0.859	1.000	1.000

Overall, the bootstrap test has no size distortions and produced good powers for moderate sample sizes.

GARCH(1,1) model

$$\begin{aligned} X_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned} \tag{6}$$

where ϵ_t are iid r.v.s with $E\epsilon_1 = 0$, $E\epsilon_1^2 = 1$ and $E\epsilon_1^4 < \infty$ and $\theta = (\omega, \alpha, \beta)'$ with $\omega > 0, \alpha \geq 0, \beta \geq 0$, and $\alpha + \beta < 1$ are in a compact subset of R^3.

- $\{X_t\}$ is ergodic and stationary (cf. Bougerol and Picard (1992 a,b))

Gaussianity test

H_0 : $\epsilon_t \sim F_0$ vs. H_1 : not H_0, where F_0 denotes a standard normal distribution function.

- $\hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)'$ of θ, $\sqrt{n}(\hat{\theta}_n - \theta) = O_P(1)$

  (e.g., QMLE in Francq and Zakoïan (2004))

- $\hat{\epsilon}_t = X_t / \hat{\sigma}_t$,

  $\hat{\sigma}_t^2 = \hat{\omega}_n + \hat{\alpha}_n X_{t-1}^2 + \hat{\beta}_n \sigma_{t-1}^2$

- $\hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s), \; 0 \le s \le 1,$

  $\hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t) \le s)$

- $\hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)'$ of θ, $\sqrt{n}(\hat{\theta}_n - \theta) = O_P(1)$

  (e.g., QMLE in Francq and Zakoïan (2004))

- $\hat{\epsilon}_t = X_t / \hat{\sigma}_t$,

  $\hat{\sigma}_t^2 = \hat{\omega}_n + \hat{\alpha}_n X_{t-1}^2 + \hat{\beta}_n \sigma_{t-1}^2$

- $\hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s),\ 0 \le s \le 1,$

  $\hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t) \le s)$

- $\hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)'$ of θ, $\sqrt{n}(\hat{\theta}_n - \theta) = O_P(1)$

  (e.g., QMLE in Francq and Zakoïan (2004))

- $\hat{\epsilon}_t = X_t / \hat{\sigma}_t$,

  $\hat{\sigma}_t^2 = \hat{\omega}_n + \hat{\alpha}_n X_{t-1}^2 + \hat{\beta}_n \sigma_{t-1}^2$

- $\hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s), \ 0 \leq s \leq 1,$

  $\hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t) \leq s)$

According to Theorem 1 of Escanciano (2010),

$$\hat{\mathcal{V}}_n(s) \quad = \quad \mathcal{V}_n(s) + R_n(s)$$

where $\mathcal{V}_n(s) = \sqrt{n}(F_n(s) - s)$, $F_n(s) = \frac{1}{n}\sum_{t=1}^{n} I(F_0(\epsilon_t) \leq s)$,

$f_0 = F_0^{'}$,and $R_n(s) = \sqrt{n}(\hat{\theta}_n - \theta_0)' a F_0^{-1}(s) f_0(F_0^{-1}(s)) + o_P(1)$

uniformly in s for some vector a.

Theorem

Under H_0, $\max_{1\le i\le m}|s_i - s_{i-1}| \to 0$ *and* $m \to \infty$, *as* $n \to \infty$,

$$\sqrt{n}\ \sup_{w\in W}|S^w(\hat{F}_n)| \overset{d}{\approx} \sup_{w\in W}\left|\sum_{i=1}^{m} w_i\left(\mathcal{BB}(s_i) - \mathcal{BB}(s_{i-1})\right)\right|,$$

where W *denotes the space of bounded weight functions* $w : [0,1] \to [0,1]$ *with* $\sum_{i=1}^m w_i = 1$, *and* $0 = s_0 \le s_1 \le \ldots \le s_m = 1$.

The test can be extended to: Escanciano's(2010)

- heterosckedastic models such as ARMA-GARCH models and threshold GARCH models.
- GARCH models with common fat-tailed error distributions.

The test can be extended to: Escanciano's(2010)

- heterosckedastic models such as ARMA-GARCH models and threshold GARCH models.

- GARCH models with common fat-tailed error distributions.

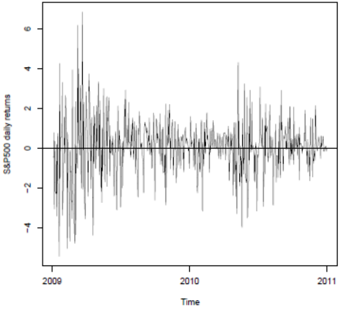
The test can be extended to: Escanciano's(2010)

- heterosckedastic models such as ARMA-GARCH models and threshold GARCH models.
- GARCH models with common fat-tailed error distributions.

Application to real data

the daily log return of S&P500 index from JAN2009 to DEC2010, $n = 503$

1. Gaussianity test:



- GARCH(1,1) model is fitted

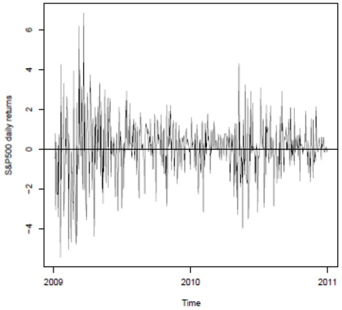
  - the QMLE for $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ = (0.017, 0.092, 0.898)

- performed with $B = 500$, $m = 7$
- rejected H_0 at 0.01, 0.05,0.1

Application to real data

the daily log return of S&P500 index from JAN2009 to DEC2010, $n = 503$

1. Gaussianity test:

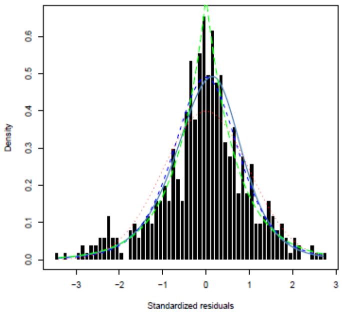


- GARCH(1,1) model is fitted

  - the QMLE for $(\hat{\omega}, \hat{\alpha}, \hat{\beta})$ = (0.017, 0.092, 0.898)

- performed with $B = 500$, $m = 7$
- rejected H_0 at 0.01, 0.05,0.1

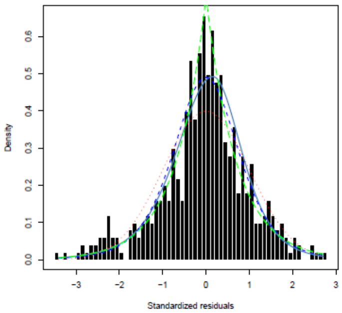
2.Tests for fat-tailed distributions:



- H_0:
  - $t(5)$
  - skewed-$t(5)$ with λ=0.9,
  - generalized error distribution with λ=1.0
- did not reject all of H_0's,
- the generalized error distribution is best fitted.

Extension of the test to common fat-tailed distribution is very practical.

2.Tests for fat-tailed distributions:



- H_0:
  - $t(5)$
  - skewed-$t(5)$ with λ=0.9,
  - generalized error distribution with λ=1.0
- did not reject all of H_0's,
- the generalized error distribution is best fitted.

Extension of the test to common fat-tailed distribution is very practical.

Summary

- Entropy-based gof test for time series models.

  - Asymptotic distribution

  - Bootstrap method

  - Validity of the test

- Covered broad class linear models and nonlinear models.

  - AR models include unstable non-stationary models

  - GARCH(1,1) models

Summary

- Entropy-based gof test for time series models.

  - Asymptotic distribution

  - Bootstrap method

  - Validity of the test

- Covered broad class linear models and nonlinear models.

  - AR models include unstable non-stationary models

  - GARCH(1,1) models

Thank you!

Simulation study

Generate GARCH(1,1) processes,

$$\begin{aligned} X_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

$\epsilon_t \sim$ iid $N(0,1)$

$\epsilon_t \sim$ iid $(1-p)\, N(\mu_1, \sigma_1^2) + p\, N(\mu_2, \sigma_2^2)$

$\theta = (\omega, \alpha, \beta)$

$(p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$

the empirical sizes

$\epsilon_t \sim$ iid $N(0,1)$

$\theta = (\omega, \alpha, \beta)$	$n = 300$, $m = 6$	$n = 400$, $m = 7$	$n = 500$, $m = 7$	$n = 1000$, $m = 10$
$(0.2, 0.2, 0.2)$	0.051	0.037	0.050	0.052
$(0.2, 0.2, 0.4)$	0.045	0.041	0.048	0.048
$(0.2, 0.2, 0.7)$	0.053	0.045	0.045	0.049
$(0.2, 0.1, 0.8)$	0.048	0.046	0.050	0.048

the empirical powers

$\epsilon_t \sim$ iid $(1-p)\, N(\mu_1, \sigma_1^2) + p\, N(\mu_2, \sigma_2^2)$

$(p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = (0.2, 0.0, 0.0, 0.75, 2.0)$

$\theta = (\omega, \alpha, \beta)$	$n = 300$ $m = 6$	$n = 400$ $m = 7$	$n = 500$ $m = 7$	$n = 1000$ $m = 10$
$(0.2, 0.2, 0.2)$	0.735	0.882	0.994	1.000
$(0.2, 0.2, 0.4)$	0.762	0.879	0.993	1.000
$(0.2, 0.2, 0.7)$	0.817	0.930	0.933	1.000
$(0.2, 0.1, 0.8)$	0.758	0.900	0.995	1.000

the empirical powers

$$\epsilon_t \sim \text{iid } (1-p)\,N(\mu_1, \sigma_1^2) + p\,N(\mu_2, \sigma_2^2)$$

$$(p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = (0.2, 0.4, -0.1, 0.7, 2.0)$$

$\theta = (\omega, \alpha, \beta)$	$n = 300$ $m = 6$	$n = 400$ $m = 7$	$n = 500$ $m = 7$	$n = 1000$ $m = 10$
$(0.2, 0.2, 0.2)$	0.844	0.990	0.990	1.000
$(0.2, 0.2, 0.4)$	0.896	0.975	0.995	1.000
$(0.2, 0.2, 0.7)$	0.920	0.975	0.995	1.000
$(0.2, 0.1, 0.8)$	0.957	0.975	0.979	1.000

the empirical powers

$$\epsilon_t \sim \text{iid } (1-p)\, N(\mu_1, \sigma_1^2) + p\, N(\mu_2, \sigma_2^2)$$

$$(p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = (0.1, 0.0, 0.0, 1.0, 10.0)$$

$\theta = (\omega, \alpha, \beta)$	$n = 300$ $m = 6$	$n = 400$ $m = 7$	$n = 500$ $m = 7$	$n = 1000$ $m = 10$
$(0.2, 0.2, 0.2)$	0.954	0.960	0.995	1.000
$(0.2, 0.2, 0.4)$	0.967	0.972	0.990	1.000
$(0.2, 0.2, 0.7)$	0.938	0.952	0.990	1.000
$(0.2, 0.1, 0.8)$	0.976	0.920	0.980	1.000

Ref. Entropy-based test of fit

- Maximum entropy test for autoregressive models. S. Lee and S. Park (2013) *Uncertainty analysis in econometrics with applications.* 200 119-128.
- Maximum entropy test for GARCH models. S. Lee, J. Lee, and S. Park (2013) *submitted*
- Maximum entropy type test of fit. S. Lee et al. (2010) *Comput. Statis. Data Anal.* 55 2635-2643.
- Maximum entropy type test of fit: Composite hypothesis case S. Lee(2013) *Comput. Statis. Data Anal.* 57 59-67.

Entropy

- Jaynes, E.T., 1957. Information theory and statistical mechanics. *Phys. Rev.* 106, 620-630.
- Forte, B., Hughes, W., 1988. The maximum entropy principle: a tool to define new entropies. *Rep. Math. Phys.* 26, 227-235.

Empirical process

- Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* 1, 279-290.
- Stute, W., Manteiga, W. G. and Quindimil, M. P. (1993). Bootstrap based goodness-of-Fit tests. *Metrika*, 40 243-256.
- Lee, S. and Wei, C. Z. (1999). On residual empirical processes of stochastic regression models with applications to time series. *Ann. Statist.* 27, 237-261.
- Escaciano, J. C. (2010). Asymptotic distribution-free diagnostic tests for heteroskedastic time series models. *Econometric Theory.* 26 744-773.

Models

- Francq, C. and Zako¨an, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605-637.
- Chan, N. H. andWei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist* 16, 367-401.
- Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and some nonnegative time series. *J. Econometrics*, 52(1-2), 115-127.
- Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. it Ann. Probab., 20(4), 1714-1730.

Normality test

- Bera, A. and Jarque, C. (1981). Efficient tests for normality, heteroskedasticity, andserial independence of regression residuals: monte carlo evidence. *Econometric Letters.* 7, 313-318.
- A note on the Jarque-Bera normality test for GARCH models. S. Lee, T. Lee, and S. Park(2010) *J. Korean Statist. Soc.* 39 93-102.

Generalized error distribution

Nelson(91) proposed to use GED to capture the fat tails observed in financial time series

$$f(\epsilon_t) = \frac{\nu exp[-(1/2)|\epsilon_t/\lambda|^{\nu}]}{\lambda \cdot 2^{(\nu+1)/\nu}\Gamma(1/\nu)}$$

where $\lambda = [\frac{2^{-2/\nu}\Gamma(1/\nu)}{\Gamma(3/\nu)}]^{1/2}$ and $\Gamma(\cdot)$ is the gamma function.

$\nu = 2$: the standard normal pdf, $\nu > 2$: thinner tails then the normal pdf, $\nu < 2$: thicker tails then the normal pdf.

$\nu = 1$: the double exponential pdf.

student t-distribution

$$f(\epsilon_t) = \frac{\Gamma[(\nu+1)/2]}{(\pi\nu)^{1/2}\Gamma(\nu/2)} \frac{\lambda^{-1/2}}{[1+\epsilon_t^2/(\lambda\nu)]^{(\nu+1)/2}}$$

where ν is degrees of freedom and λ is the scale parameter

the scale parameter λ should be chosen to be $\frac{\sigma^2(\nu-2)}{\nu}$.

skewed t-distribution

$$f(\epsilon_t) = \frac{\Gamma[(\nu+k)/2]}{(\pi\nu)^{k/2}\Gamma(\nu/2)} \frac{\lambda^{-1/2}}{[1+\epsilon_t^2/(\lambda\nu)]^{(\nu+k)/2}}$$

where ν is degrees of freedom and k >0

the scale parameter λ should be chosen to be $\frac{\sigma^2(\nu-2)}{\nu}$.