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.
Boltzmann-Shannon entropy:

\[ H(f) = - \int_{-\infty}^{\infty} f(x) \log(f(x)) \, dx. \] (1)

Forte and Hughes (1988) proposed the function

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

as the discrete analogue of (1), \( p_i = \int_{x_{i-1}}^{x_i} f(x) \, dx \).
for a variable defined in \([a, b]\),

\[
\lim_{\max_i |x_i - x_{i-1}| \to 0} \left( - \sum_{i=1}^{n} p_i \log\left(\frac{p_i}{(x_i - x_{i-1})}\right) \right) = H(f),
\]

(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 $F,$

$$H_0 : F = F_0 \quad \text{vs.} \quad F \neq F_0.$$
A generalization of Forte and Hughes (1988)’s:

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

(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.
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 $n \to \infty$,

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

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,

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

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

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

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

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

It can be seen that

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

Hence, the previous theorem can be applied when

\[ \max_{1 \leq i \leq m} |s_i - s_{i-1}| \to 0 \quad \text{and} \quad m \to \infty. \]
We can express
\[ \hat{E}_n(s) = E_n(s) - h(s)' \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1), \]
where \( h(s) = \frac{\partial F_{\theta_0}(F_{\theta_0}^{-1}(s))}{\partial \theta} \).

It can be seen that
\[
\sqrt{n}S^w(\hat{F}_n) = \left| \sum_{i=1}^{m} w_i [E_n(s_i) - E_n(s_{i-1})] \right|
+ \sqrt{n}(\hat{\theta}_n - \theta_0)' \sum_{i=1}^{m} w_i [h(s_{i-1}) - h(s_i)] + o_P(1).
\]

Hence, the previous theorem can be applied when
\[ \max_{1 \leq i \leq m} |s_i - s_{i-1}| \to 0 \text{ and } m \to \infty. \]
We can express

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

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

It can be seen that

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

Hence, the previous theorem can be applied when

$$\max_{1 \leq i \leq m} |s_i - s_{i-1}| \to 0 \text{ and } m \to \infty.$$
Theorem

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

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

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 $0 = s_0 \leq s_1 \leq \ldots \leq s_m = 1$. 
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 $L \to \infty$,

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

- $L=1000$ 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 $L \to \infty$,

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

- $L=1000$ 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 $L \to \infty$,

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

- $L=1000$ is recommended.
Practical issues: $\sup_w$

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

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

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

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

- $L=1000$ 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:

$$\hat{T}_n = \sqrt{n} \max_{1 \leq l \leq L} \left| \sum_{i=1}^{m} w_{li}(\hat{F}_n(i/m) - \hat{F}_n((i - 1)/m)) \times \log m(\hat{F}_n(i/m) - \hat{F}_n((i - 1)/m)) \right|.$$
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 \( m/n \to 0. \) \( m \approx n^{1/3} \) is recommended.

Test statistic:

\[
\hat{T}_n = \sqrt{n} \max_{1 \leq l \leq L} \left| \sum_{i=1}^{m} w_{li} (\hat{F}_n(i/m) - \hat{F}_n((i - 1)/m)) \right| \times \log m(\hat{F}_n(i/m) - \hat{F}_n((i - 1)/m)).
\]
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 $m/n \rightarrow 0$. $m \approx n^{1/3}$ is recommended.

Test statistic:

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

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

\[ X_t - \beta_1 X_{t-1} - \cdots - \beta_q X_{t-q} = \epsilon_t, \]  

where \( \epsilon_t \) are iid r.v.s with \( E\epsilon_1 = 0, E\epsilon_1^2 = \sigma^2 \), and \( E\epsilon_1^4 < \infty \).

\[ \phi(z) = 1 - \beta_1 z - \cdots - \beta_q z^q \]

\[ = (1 - z)^a (1 + z)^b \prod_{k=1}^l (1 - 2 \cos \theta_k z + z^2)^{d_k} \psi(z), \]

where \( a, b, l, d_k \) are nonnegative integers, \( \theta_k \) 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 \( \epsilon_t \) are iid r.v.s with \( E\epsilon_1 = 0 \), \( E\epsilon_1^2 = \sigma^2 \), and \( E\epsilon_1^4 < \infty \).

\[
\phi(z) = 1 - \beta_1 z - \cdots - \beta_q z^q \\
= (1 - z)^a (1 + z)^b \prod_{k=1}^{l} (1 - 2 \cos \theta_k z + z^2)^{d_k} \psi(z),
\]

where \( a, b, l, d_k \) are nonnegative integers, \( \theta_k \) belongs to \((0, \pi)\) and \( \psi(z) \) is the polynomial of order \( r = q - (a + b + 2d_1 + \cdots + 2d_l) \), 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 \( \epsilon_t \) are iid r.v.s with \( E\epsilon_1 = 0, E\epsilon_1^2 = \sigma^2 \), and \( E\epsilon_1^4 < \infty \).

\[
\phi(z) = 1 - \beta_1 z - \cdots - \beta_q z^q
\]

\[
= (1 - z)^a (1 + z)^b \prod_{k=1}^{l} (1 - 2 \cos \theta_k z + z^2)^{d_k} \psi(z),
\]

where \( a, b, l, d_k \) are nonnegative integers, \( \theta_k \) belongs to \((0, \pi)\) and \( \psi(z) \) is the polynomial of order \( r = q - (a + b + 2d_1 + \cdots + 2d_l) \) 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 $\hat{\beta}_n$ of $\beta_0 = (\beta_1, \ldots, \beta_q)'$

- 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 $\hat{\epsilon}_t = X_t - \hat{\beta}_n'X_{t-1}$ and $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2$. 
LSE $\hat{\beta}_n$ of $\beta_0 = (\beta_1, \ldots, \beta_q)'$

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

the residual empirical process:

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

where $\hat{\epsilon}_t = X_t - \hat{\beta}_n'X_{t-1}$ and $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^{n} \hat{\epsilon}_t^2$. 
Lee and Wei (1999)

\[
\hat{\mathcal{V}}_n(s) = \mathcal{V}_n(s) - (\hat{\beta}_n - \beta_0)' \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{t-1} f_0(F_0^{-1}(s)) \\
+ \frac{\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)}{2\sigma_0^2} f_0(F_0^{-1}(s))F_0^{-1}(s) + \Delta_n(s),
\]

where \(\sigma_0\) and \(\beta_0\) 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
$max_{1 \leq i \leq m} |s_i - s_{i-1}| \to 0$ and $m \to \infty$, as $n \to \infty$, under $H_0$,

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

where $W$ denotes the space of bounded weight functions $w : [0, 1] \to [0, 1]$ with $\sum_{i=1}^m w_i = 1$, and $0 = s_0 \leq s_1 \leq \ldots \leq s_m = 1$. 
Bootstrap method: Stute et al. (1993)

**Step 1:** Obtain the LSE $\hat{\beta}_n$ and $\hat{\sigma}^2_n$ 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 Step 3.
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 Step 3.
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,

\[ X_t = \beta X_{t-1} + \epsilon_t, \; \epsilon_t \sim \text{iid } N(0, 1) \]

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

for \( \beta = 0.1, 0.3, 0.5, 1.0, \; p = 0.9, \; \text{and } \sigma_1^2 = 10, 25 \)

Replicate 1000 times with \( B = 500 \) at the nominal level 0.05
Simulation study

Generate AR(1) processes,

\[ X_t = \beta X_{t-1} + \epsilon_t, \quad \epsilon_t \sim \text{iid } N(0, 1) \]

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

for \( \beta = 0.1, 0.3, 0.5, 1.0, \ p = 0.9, \ \text{and } \sigma_1^2 = 10, 25 \)

Replicate 1000 times with \( B = 500 \) at the nominal level 0.05
the empirical sizes

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 3$</td>
<td>$m = 5$</td>
<td>$m = 7$</td>
</tr>
<tr>
<td>0.3</td>
<td>0.049</td>
<td>0.056</td>
<td>0.053</td>
</tr>
<tr>
<td>0.5</td>
<td>0.046</td>
<td>0.043</td>
<td>0.057</td>
</tr>
<tr>
<td>0.7</td>
<td>0.043</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>1.0</td>
<td>0.051</td>
<td>0.048</td>
<td>0.044</td>
</tr>
</tbody>
</table>

- the test has no size distortions.

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

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 3$</td>
<td>$m = 5$</td>
<td>$m = 7$</td>
</tr>
<tr>
<td>0.3</td>
<td>0.049</td>
<td>0.056</td>
<td>0.053</td>
</tr>
<tr>
<td>0.5</td>
<td>0.046</td>
<td>0.043</td>
<td>0.057</td>
</tr>
<tr>
<td>0.7</td>
<td>0.043</td>
<td>0.051</td>
<td>0.051</td>
</tr>
<tr>
<td>1.0</td>
<td>0.051</td>
<td>0.048</td>
<td>0.044</td>
</tr>
</tbody>
</table>

- the test has no size distortions.
- also stable size for large $n(\geq 500)$. 
the empirical powers

<table>
<thead>
<tr>
<th>$\sigma^2_1$</th>
<th>$\beta$</th>
<th>$n = 100$</th>
<th>$n = 300$</th>
<th>$n = 500$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$m = 3$</td>
<td>$m = 5$</td>
<td>$m = 7$</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.3</td>
<td>0.386</td>
<td>0.771</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.368</td>
<td>0.767</td>
<td>0.928</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.394</td>
<td>0.747</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.410</td>
<td>0.787</td>
<td>0.954</td>
</tr>
<tr>
<td>25</td>
<td>0.3</td>
<td>0.823</td>
<td>0.997</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.828</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.852</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>1.0</td>
<td>0.859</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Overall, the bootstrap test has no size distortions and produced good powers for moderate sample sizes.
GARCH(1,1) model

\[ X_t = \sigma_t \epsilon_t, \]
\[ \sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \] (6)

where \( \epsilon_t \) 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

\[ H_0 : \epsilon_t \sim F_0 \text{ vs. } H_1 : \text{not } H_0, \text{ where } F_0 \text{ denotes a standard normal distribution function.} \]
\[ \hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)' \text{ of } \theta, \sqrt{n}(\hat{\theta}_n - \theta) = O_P(1) \]

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

\[ \hat{\epsilon}_t = X_t/\hat{\sigma}_t, \]
\[ \hat{\sigma}_t^2 = \hat{\omega}_n + \hat{\alpha}_n X_{t-1}^2 + \hat{\beta}_n \sigma_{t-1}^2 \]

\[ \hat{\mathcal{V}}_n(s) = \sqrt{n}(\hat{F}_n(s) - s), \quad 0 \leq s \leq 1, \]
\[ \hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t) \leq s) \]
\[ \hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n)' \text{ of } \theta, \quad \sqrt{n}(\hat{\theta}_n - \theta) = O_P(1) \]

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

\[ \hat{\epsilon}_t = X_t / \hat{\sigma}_t, \]

\[ \hat{\sigma}_t^2 = \hat{\omega}_n + \hat{\alpha}_nX_{t-1}^2 + \hat{\beta}_n\sigma_{t-1}^2 \]

\[ \hat{\nu}_n(s) = \sqrt{n}(\hat{F}_n(s) - s), \quad 0 \leq s \leq 1, \]

\[ \hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t) \leq s) \]
\[ \hat{\theta}_n = (\hat{\omega}_n, \hat{\alpha}_n, \hat{\beta}_n) \] of \( \theta \), \( \sqrt{n}(\hat{\theta}_n - \theta) = O_P(1) \)

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

\[ \hat{\epsilon}_t = X_t / \hat{\sigma}_t, \]
\[ \hat{\sigma}_t^2 = \hat{\omega}_n + \hat{\alpha}_n X_{t-1}^2 + \hat{\beta}_n \sigma_{t-1}^2 \]

\[ \hat{V}_n(s) = \sqrt{n}(\hat{F}_n(s) - s), \quad 0 \leq s \leq 1, \]
\[ \hat{F}_n(s) = \frac{1}{n} \sum_{t=1}^{n} I(F_0(\hat{\epsilon}_t) \leq s) \]
According to Theorem 1 of Escanciano (2010),

\[ \hat{\mathcal{V}}_n(s) = \mathcal{V}_n(s) + R_n(s) \]

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 $H_0$, \( \max_{1 \leq i \leq m} |s_i - s_{i-1}| \to 0 \) and \( m \to \infty \), as \( n \to \infty \),

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

where \( W \) denotes the space of bounded weight functions \( w : [0, 1] \to [0, 1] \) with \( \sum_{i=1}^{m} w_i = 1 \), and

\( 0 = s_0 \leq s_1 \leq \ldots \leq s_m = 1 \).
The test can be extended to: Escanciano’s (2010)

- heteroscedastic 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)

- heteroscedastic 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)

- heteroskedastic 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
Application to real data

2. Tests for fat-tailed distributions:

- $H_0$:
  - $t(5)$
  - skewed-$t(5)$ with $\lambda=0.9$
  - generalized error distribution with $\lambda=1.0$

- 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.
Application to real data

2. Tests for fat-tailed distributions:

- $H_0$:
  - $t(5)$
  - skewed-$t(5)$ with $\lambda=0.9$
  - generalized error distribution with $\lambda=1.0$

- 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,

\[
X_t = \sigma_t \epsilon_t, \\
\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2,
\]

\(\epsilon_t \sim \text{iid } N(0, 1)\)

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

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

\((p, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)\)
the empirical sizes

\[ \epsilon_t \sim \text{iid } N(0, 1) \]

<table>
<thead>
<tr>
<th>( \theta = (\omega, \alpha, \beta) )</th>
<th>( n = 300 )</th>
<th>( n = 400 )</th>
<th>( n = 500 )</th>
<th>( n = 1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (0.2, 0.2, 0.2) )</td>
<td>0.051</td>
<td>0.037</td>
<td>0.050</td>
<td>0.052</td>
</tr>
<tr>
<td>( (0.2, 0.2, 0.4) )</td>
<td>0.045</td>
<td>0.041</td>
<td>0.048</td>
<td>0.048</td>
</tr>
<tr>
<td>( (0.2, 0.2, 0.7) )</td>
<td>0.053</td>
<td>0.045</td>
<td>0.045</td>
<td>0.049</td>
</tr>
<tr>
<td>( (0.2, 0.1, 0.8) )</td>
<td>0.048</td>
<td>0.046</td>
<td>0.050</td>
<td>0.048</td>
</tr>
</tbody>
</table>
the empirical powers

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

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

<table>
<thead>
<tr>
<th>(\theta = (\omega, \alpha, \beta))</th>
<th>(n = 300)</th>
<th>(n = 400)</th>
<th>(n = 500)</th>
<th>(n = 1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((0.2, 0.2, 0.2))</td>
<td>0.735</td>
<td>0.882</td>
<td>0.994</td>
<td>1.000</td>
</tr>
<tr>
<td>((0.2, 0.2, 0.4))</td>
<td>0.762</td>
<td>0.879</td>
<td>0.993</td>
<td>1.000</td>
</tr>
<tr>
<td>((0.2, 0.2, 0.7))</td>
<td>0.817</td>
<td>0.930</td>
<td>0.933</td>
<td>1.000</td>
</tr>
<tr>
<td>((0.2, 0.1, 0.8))</td>
<td>0.758</td>
<td>0.900</td>
<td>0.995</td>
<td>1.000</td>
</tr>
</tbody>
</table>
the empirical powers

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

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

<table>
<thead>
<tr>
<th>( \theta = (\omega, \alpha, \beta) )</th>
<th>( n = 300 )</th>
<th>( n = 400 )</th>
<th>( n = 500 )</th>
<th>( n = 1000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.2, 0.2, 0.2)</td>
<td>0.844</td>
<td>0.990</td>
<td>0.990</td>
<td>1.000</td>
</tr>
<tr>
<td>(0.2, 0.2, 0.4)</td>
<td>0.896</td>
<td>0.975</td>
<td>0.995</td>
<td>1.000</td>
</tr>
<tr>
<td>(0.2, 0.2, 0.7)</td>
<td>0.920</td>
<td>0.975</td>
<td>0.995</td>
<td>1.000</td>
</tr>
<tr>
<td>(0.2, 0.1, 0.8)</td>
<td>0.957</td>
<td>0.975</td>
<td>0.979</td>
<td>1.000</td>
</tr>
</tbody>
</table>
the empirical powers

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

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

<table>
<thead>
<tr>
<th>\theta = (\omega, \alpha, \beta)</th>
<th>\theta = (\omega, \alpha, \beta)</th>
<th>\theta = (\omega, \alpha, \beta)</th>
<th>\theta = (\omega, \alpha, \beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\theta = (0.2, 0.2, 0.2)</td>
<td>0.954</td>
<td>0.960</td>
<td>0.995</td>
</tr>
<tr>
<td>\theta = (0.2, 0.2, 0.4)</td>
<td>0.967</td>
<td>0.972</td>
<td>0.990</td>
</tr>
<tr>
<td>\theta = (0.2, 0.2, 0.7)</td>
<td>0.938</td>
<td>0.952</td>
<td>0.990</td>
</tr>
<tr>
<td>\theta = (0.2, 0.1, 0.8)</td>
<td>0.976</td>
<td>0.920</td>
<td>0.980</td>
</tr>
</tbody>
</table>
Ref. Entropy-based test of fit


- Maximum entropy test for GARCH models. S. Lee, J. Lee, and S. Park (2013) *submitted*


Entropy

Empirical process


Models


Normality test


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 \( \lambda = \left[\frac{2^{-2/\nu} \Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2} \) 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 \( \nu \) is degrees of freedom and \( \lambda \) is the scale parameter. 

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

S.Lee, J. Lee, and S. Park

ENTROPY-BASED TEST FOR TIME SERIES MODELS
A skewed t-distribution is given by:

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

where $\nu$ is degrees of freedom and $k > 0$.

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