DOCTORAL PROGRAM
SWISS FINANCE INSTITUTE

FINANCIAL ECONOMETRICS

swiss:finance:institute

Eric Jondeau

FINANCIAL ECONOMETRICS

Lecture 3: Modeling Non-Normality

Eric Jondeau

Objectives of the lecture

The objective of this session is the **modeling of the conditional distribution**.

Distinction conditional vs. unconditional distribution

- Time-varying volatility generates fat tails in the unconditional distribution, but this is not enough to fit the actual data.
- We need to introduce a non-normal conditional distribution to adjust the data.

We focus on two particular stylized facts of actual data:

- The asymmetry and
- The fat-tailedness of the empirical distribution.

How to extend the normal distribution into these two directions?

- There exist natural extensions to the normal distribution allowing fat tails (such as the Student *t* distribution)
- But these distributions are not designed to capture asymmetry.

Objectives of the lecture

➔ **Non-normal Distributions**

- **QMLE**
- **Existence of moments**
- **Modeling the conditional distribution**
- **Adequacy tests**

The model

Let r_t, for $t = 1, \cdots, T$, be a time series of realizations of log-returns.

Then, it is convenient to break down the dynamics of r_t in three components:

(1) **the conditional mean** which contains location parameters (mean, median, or mode) and specifies the locations of the range of values

(2) **the conditional variance** which contains scale parameters and measures the variability of the pdf; and finally

(3) **the conditional distribution** which contains shape parameters and determines the form of a distribution within the general family of distributions.

The model

The model can be written as

$$r_t = \mu_t(\theta) + \varepsilon_t \qquad \text{with} \qquad \mu_t(\theta) = E[r_t \mid I_{t-1}] = \mu(\theta, I_{t-1})$$
$$\varepsilon_t = \sigma_t(\theta) z_t \qquad \text{with} \qquad \sigma_t^2(\theta) = E[(r_t - \mu_t(\theta))^2 \mid I_{t-1}] = \sigma^2(\theta, I_{t-1})$$
$$z_t \sim g(z_t \mid \eta)$$

The innovation, $z_t = (r_t - \mu_t(\theta)) / \sigma_t(\theta)$, has zero mean and unit variance.

The dynamics of volatility may be a GARCH model.

θ includes all parameters associated with the conditional mean and variance equations.

The last equation specifies the conditional distribution $g(.)$ with **shape parameters** η. When the conditional distribution is $N(0,1)$, there is no shape parameter. In the general case, η includes parameters describing the distribution.

The model

We address several issues related to the modeling of non-normal returns:

1. if the unconditional distribution has been found to be non-normal, do we necessarily have to assume that the conditional distribution $g(.)$ is non-normal?

2. if the conditional distribution is assumed to be non-normal, do we necessarily have to model it explicitly?

3. if we need an explicit model for the conditional distribution, how far can we go in terms of asymmetry and fat tailedness of the distribution?

Higher moments of a GARCH process

Attractive feature of GARCH models: even when the conditional distribution of innovations z_t is normal, the unconditional distribution of the error term ε_t (and of r_t) has fatter tails than the normal distribution. (Engle and González-Rivera, 1991, *JBES*).

We distinguish between unconditional and conditional kurtosis

$$K_u = \frac{E[\varepsilon_t^4]}{(E[\sigma_t^2])^2} = \frac{E[\varepsilon_t^4]}{\sigma^4} \qquad \text{and} \qquad K_c = \frac{E[\varepsilon_t^4 \mid I_{t-1}]}{\sigma_t^4}$$

The **unconditional kurtosis** is given by, assuming $E[\varepsilon_t^4] < \infty$,

$$K_u = \frac{E[\varepsilon_t^4]}{(E[\sigma_t^2])^2} = \frac{E[E[\varepsilon_t^4 \mid I_{t-1}]]}{(E[\sigma_t^2])^2} = \frac{E[K_c \sigma_t^4]}{(E[\sigma_t^2])^2} = K_c \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2}$$

We have $E[\sigma_t^4] \geq (E[\sigma_t^2])^2$ for $\sigma_t > 0$ (Jensen's inequality) so that: $K_u \geq K_c$.

Even in the case where innovations z_t are assumed to be normal, a GARCH model yields to an unconditional distribution with fatter tails than the normal distribution.

Jensen's inequality: for a convex function φ, $E[\varphi(x)] \geq \varphi(E[x])$. (Here, $\varphi(x) = x^2$)

Higher moments of a GARCH process − Skewness

Similarly, we have the unconditional and conditional skewness:

$$S_u = \frac{E[\varepsilon_t^3]}{(E[\sigma_t^2])^{3/2}} = \frac{E[\varepsilon_t^3]}{\sigma^3} \qquad \text{and} \qquad S_c = \frac{E[\varepsilon_t^3 \mid I_{t-1}]}{\sigma_t^3}$$

The **unconditional skewness** is given by, assuming $E[\varepsilon_t^3] < \infty$,

$$S_u = \frac{E[\varepsilon_t^3]}{(E[\sigma_t^2])^{3/2}} = \frac{E[E[\varepsilon_t^3 \mid I_{t-1}]]}{(E[\sigma_t^2])^{3/2}} = \frac{E[S_c \sigma_t^3]}{(E[\sigma_t^2])^{3/2}} = S_c \frac{E[\sigma_t^3]}{(E[\sigma_t^2])^{3/2}}$$

We have $E[\sigma_t^3] \geq (E[\sigma_t^2])^{3/2}$ for $\sigma_t > 0$ (Jensen's inequality) so that: $|S_u| \geq |S_c|$

Symmetric conditional distribution: introducing GARCH effects cannot result in an asymmetric distribution.

Asymmetric conditional distribution: the unconditional skewness is larger than the conditional one.

Higher moments of a GARCH process − Kurtosis

Examples

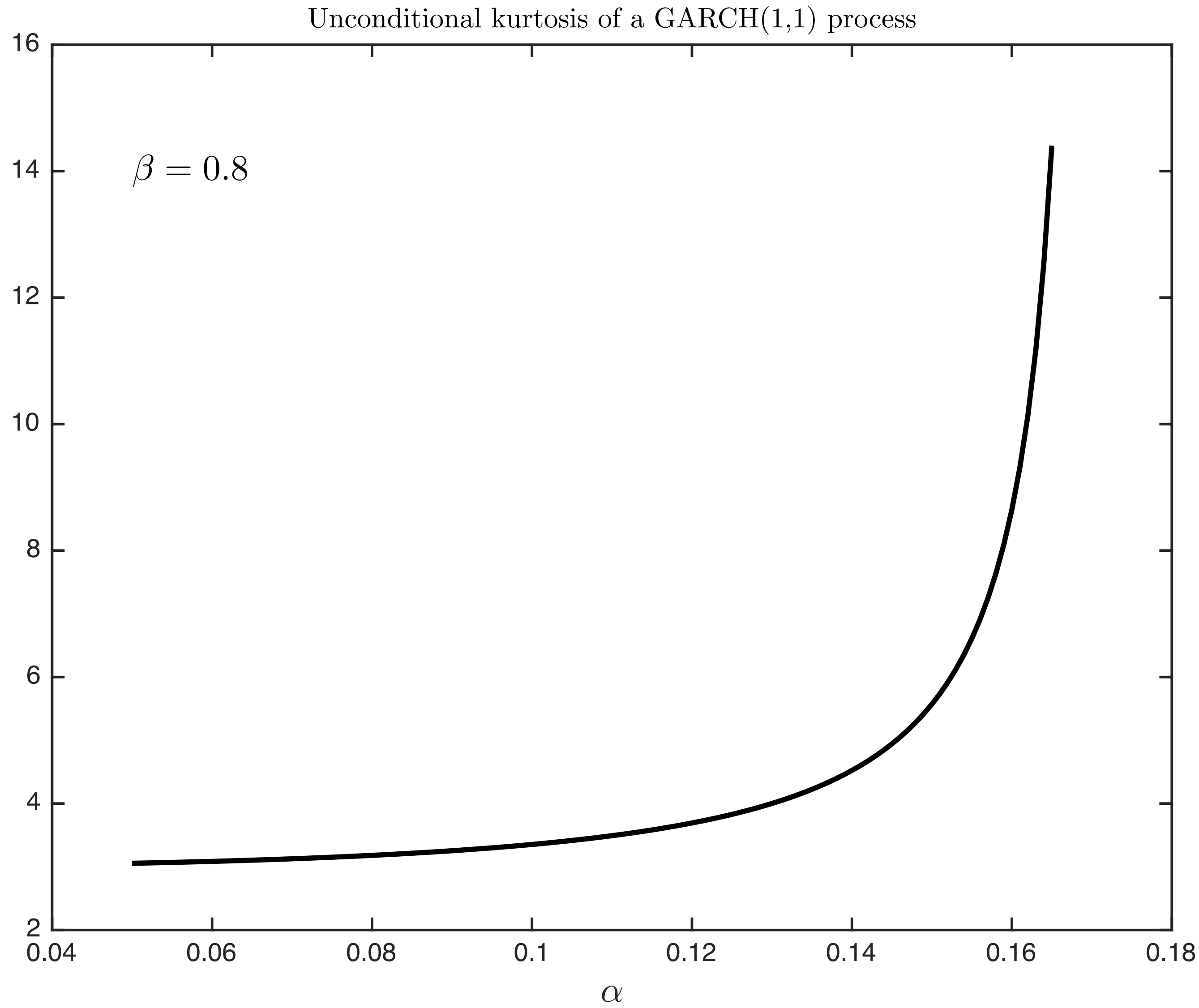
ARCH(1) model, with $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2$ and normal innovations:

$$K_u = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3 \quad \text{for } \alpha_1 > 0 \quad \text{and} \quad 3\alpha_1^2 < 1$$

GARCH(1,1) model, with $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$ and normal innovations:

$$K_u = 3\frac{1-\beta_1^2 - 2\alpha_1\beta_1 - \alpha_1^2}{1-\beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2} > 3 \quad \text{for } \alpha_1 > 0,\ \beta_1 > 0 \quad \text{and} \quad \beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1.$$

Higher moments of a GARCH process − Kurtosis



Objectives of the lecture

- Non-normal Distributions

➔ **QMLE**

- Existence of moments

- Modeling the conditional distribution

- Adequacy tests

Quasi Maximum Likelihood Estimation

If the conditional distribution g is not normal, the ML approach cannot be directly used, since it is based on the normality assumption.

Proposition: Gouriéroux, Monfort and Trognon (1984, *Econometrica*) and Bollerslev and Wooldridge (1992, *Econometric Reviews*): **If the first and second moments are correctly specified, a consistent estimate of θ is obtained by maximizing the likelihood under the assumption of conditional normality, even if the true distribution is not normal. This is the QML Estimation**.

MLE – maximizes the likelihood assuming that the true conditional distribution of errors is normal.

QMLE – maximizes the normal likelihood, even when the true distribution is non-normal.

The two estimators of θ are the same, as this is the same maximization problem. However, the covariance matrices of the estimators differ. Under the QMLE, the covariance is computed **without** assuming conditional normality.

Estimator

The **QML estimator** $\hat{\theta}_{QML}$ is obtained as solution to

$$\max_{\theta \in \Theta} \ \log(L_T(\theta \mid \varepsilon_1, \ldots, \varepsilon_T)) = \sum_{t=1}^{T} \log(\ell_t(\theta \mid \varepsilon_1, \ldots, \varepsilon_T))$$

where

$$\log(\ell_t(\theta)) = -\tfrac{1}{2}\log(\sigma_t^2) + \log(g(z_t)) = -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\log(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}$$

$$= -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}\log(\sigma_t^2) - \tfrac{1}{2} z_t^2$$

The asymptotic distribution of $\hat{\theta}_{QML}$ is

$$\sqrt{T}(\hat{\theta}_{QML} - \theta_0) \overset{a}{\sim} N(0, \Omega)$$

where θ_0 is the true value of the parameter and Ω the asymptotic covariance matrix.

Covariance matrix

The QMLE provides **"robust" standard errors** that give asymptotically valid confidence intervals for the estimators ("sandwich" estimator). They are obtained as the square roots of the diagonal elements of the matrix

$$\Omega = A_0^{-1} B_0 A_0^{-1}$$

where $A_0 = -\frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial^2 \log(\ell_t(\theta_0))}{\partial\theta\partial\theta'}\right]$ is the Hessian matrix

and $B_0 = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial \log(\ell_t(\theta_0))}{\partial\theta}\frac{\partial \log(\ell_t(\theta_0))'}{\partial\theta}\right]$ is the outer product of the gradients

evaluated at θ_0, with $\partial \log(\ell_t(\theta_0))/\partial\theta'$ the score vector of $\log(\ell_t(\theta))$.

Covariance matrix

In finite sample, the asymptotic covariance matrix is estimated by

$$\hat{\Omega}_T = \hat{A}_T^{-1} \hat{B}_T \hat{A}_T^{-1}$$

where $$\hat{A}_T = -\frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial^2 \log(\ell_t(\hat{\theta}_{QML}))}{\partial \theta \partial \theta'} \right]$$

$$\hat{B}_T = \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial \log(\ell_t(\hat{\theta}_{QML}))}{\partial \theta} \frac{\partial \log(\ell_t(\hat{\theta}_{QML}))'}{\partial \theta} \right]$$

are evaluated at the QML estimates $\hat{\theta}_{QML}$.

Remark: The asymptotic covariance matrix of the ML estimator $\hat{\theta}_{ML}$ is simply given by the inverse of the Hessian matrix so that we have

$$\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \overset{a}{\sim} N(0, A_0^{-1})$$ since, under normality, $B_0 = A_0$.

Normal QMLE: Inefficiency

If we maximize the normal likelihood: **QMLE is consistent** whatever the true distribution, provided the conditional mean and variance are correctly specified.

QMLE is **robust** with respect to the true distribution of the model.

What about efficiency?

- If the true distribution is normal, **QMLE is efficient**
- If the true distribution is non-normal, **QMLE is inefficient**

Engle and González-Rivera (1991, *JBES*) show that **the degree of inefficiency of the QMLE increases with the degree of departure from normality**.

- Simulation evidence: the loss of efficiency of normal QMLE relative to MLE under the correct distribution as high as 84% (QMLE variance is 6.25 times larger than MLE variance)
- If the true distribution is a t with 5 degrees of freedom, the loss of efficiency is 59%

Non-normal QMLE

Therefore, **maximizing the likelihood using the correct distribution of** z_t **will improve the efficiency of the estimator**.

Common practice: use **QMLE with a non-normal distribution**.

Example: Bollerslev (1987, *REStat*) estimates a GARCH model with a Student t.

In this case, the log-likelihood is maximized under the assumed distribution g:

$$\max_{(\theta,\eta)} \quad \log(L_T(\theta,\eta \mid \varepsilon_1,\ldots,\varepsilon_T)) = \sum_{t=1}^{T} \log(\ell_t(\theta,\eta \mid \varepsilon_1,\ldots,\varepsilon_T))$$

where $\quad \log(\ell_t(\theta,\eta)) = -\dfrac{1}{2}\log(\sigma_t^2(\theta)) + \log\left(g\left(\dfrac{r_t - \mu_t(\theta)}{\sigma_t(\theta)}\right) \mid \eta \right)$

We use the robustified covariance matrix of estimators (sandwich estimator) as the conditional distribution is not normal.

Non-normal QMLE: Possible Inconsistency

The **consistency of the QMLE is not guaranteed** if we use a wrong (non-normal) distribution to maximize the likelihood (Newey and Steigerwald, 1997, *Econometrica*).

If the location of the innovation distribution (median, denoted by α_0) is non-zero, then there should be an additional parameter in the conditional mean equation

$$r_t = \mu + \rho r_{t-1} + \boldsymbol{\sigma_t \alpha_0} + \sigma_t z_t \qquad \text{with} \quad \sigma_t^2 = \omega + \alpha \varepsilon_t^2 + \beta \sigma_t^2$$

Consistency of a non-normal QMLE is achieved only if either

(1) the theoretical and empirical error pdfs are unimodal and symmetric about zero; or
(2) the conditional mean of the return process is identically zero ($\mu + \rho r_{t-1} = 0$).

When these two conditions are not satisfied, the QMLE may be inconsistent, because it fails to capture the effect of the asymmetry of the distribution on the conditional mean.

This problem can be handled by introducing an additional location parameter, which makes the QMLE robust to asymmetry.

A crucial issue when a non-normal likelihood is used for the QMLE is whether adequacy tests confirm that the assumed distribution correctly fits the data.

Objectives of the lecture

- Non-normal Distributions
- QMLE

➔ **Existence of moments**

- Modeling the conditional distribution
- Adequacy tests

Existence of moments

Moment problem: Given the sequence of moments, $\{\mu_j\}$, find the necessary and sufficient conditions (nsc) for the existence of and an expression for a positive pdf.

Hamburger moment problem: The nsc is that the moment matrices (Hankel matrices of the moment sequences) are positive definite for all *n*. (case of a r.v. defined over R.)

Theorem: The moment problem has a solution. It is necessary that:

$$\det\left\|\mu_{i,j}\right\|_{i,j=0}^{n} \geq 0 \qquad \text{for } n = 0,1,\ldots$$

where $\left\|\mu_{i,j}\right\|$ denotes the Hankel matrix of order i, j.

Example of Hankel matrix: $\begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}$

Existence of moments

What are the maximal values of skewness and kurtosis which are attainable? The conditions that must be satisfied by the sequence of moments μ_j are:

$$\mu_0 \geq 0 \qquad \begin{vmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix} \geq 0 \qquad \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix} \geq 0 \qquad \ldots$$

where $\mu_i = \int z^i f(z) dz$.

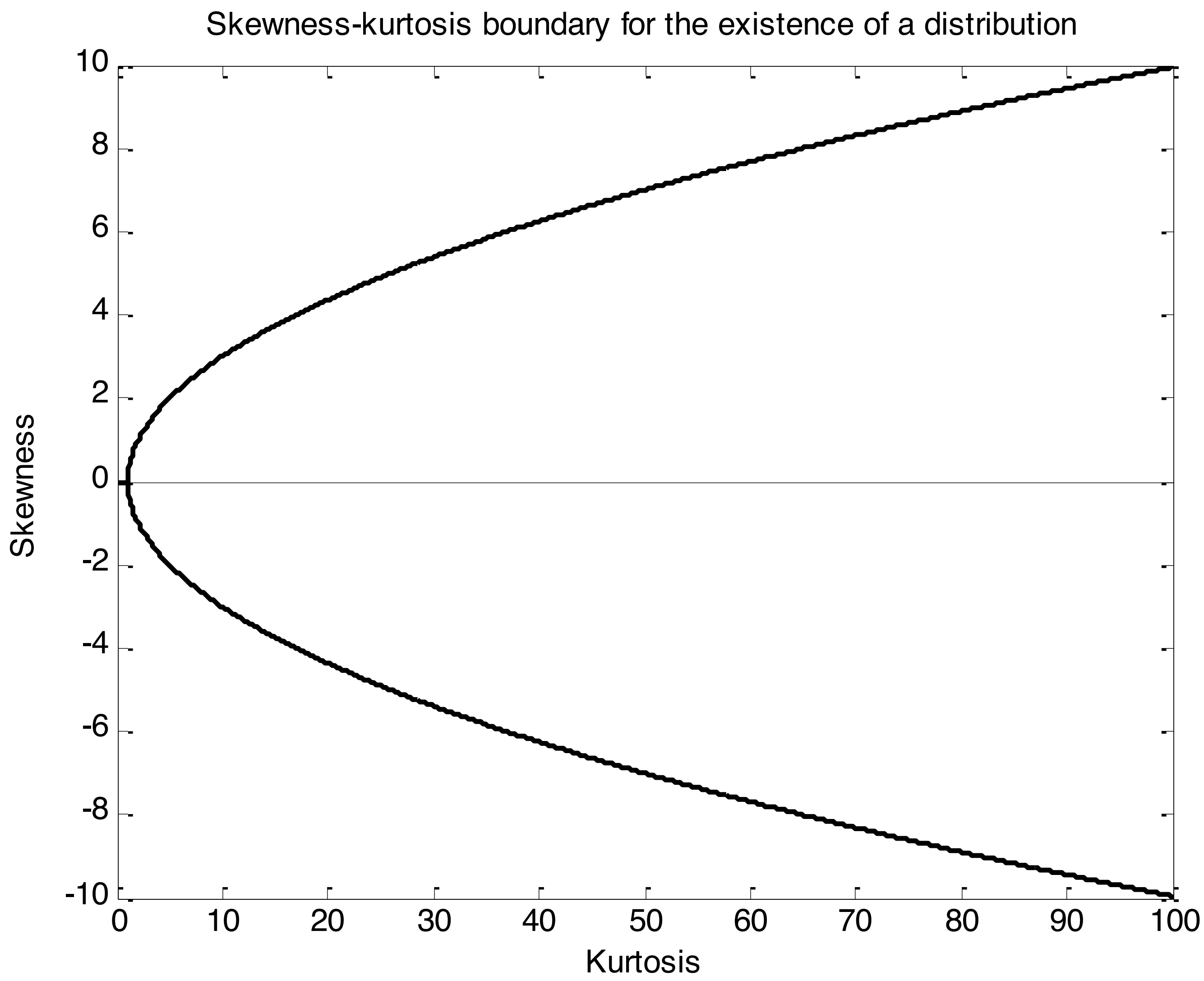
- By construction, as the pdf integrates to 1, $\mu_0 = 1$.
- As z_t is the standardized innovation, $\mu_1 = 0$, and $\mu_2 = 1$.
- So, the following relation between skewness μ_3 and kurtosis μ_4 holds:

$$\mu_3^2 \leq \mu_4 - 1 \qquad \text{with} \qquad \mu_4 \geq 1$$

For a given level of kurtosis, only a finite range of skewness can be reached.

Cases between $\mu_4 = 1$ and 3 are densities with tails thinner than the normal density

Skewness-kurtosis boundary



Objectives of the lecture

- Non-normal Distributions
- QMLE
- Existence of moments

➔ **Modeling the conditional distribution**

- Adequacy tests

Modeling the Conditional Distribution

There are several ways of dealing with asymmetry or fat tails in a distribution.

(1) A **non- or semi-parametric estimation** will capture the non-normality directly.

Engle and González-Rivera (1991, *JBES*): **Nonparametric conditional distribution**.

(2) Asymmetry can be introduced using an **expansion about a symmetric distribution**. Once the general framework has been developed, such generalization can be applied to any symmetric distribution.

Gallant, Hsieh, and Tauchen (1991): **Gram-Charlier type A expansion**.

(3) There exist some **distributions with asymmetry or fat tails**.

Hansen (1994, *IER*): **Skewed Student *t* distribution**.

Semi-parametric estimation

Semi-parametric ARCH model proposed by Engle and González-Rivera (1991):

- First and second moments are given by an ARMA process and an ARCH model.

- The conditional density is approximated by a non-parametric density estimator.

The conditional distribution $g(z_t)$ is not known, but z_t is assumed to be i.i.d. with mean 0 and variance 1. The log-likelihood function is defined by

$$\log(L_T(\psi; r_1, \ldots, r_T)) = \sum_{t=1}^{T} \log(\ell_t(\psi))$$

where

$$\log(\ell_t(\psi)) = -\frac{1}{2}\log(\sigma_t^2(\theta)) + \log(g(z_t(\theta) \mid \eta))$$

with $\psi = (\theta', \eta')'$ the vector of unknown parameters.

Semi-parametric estimation

To maximize the log-likelihood function, the following procedure is proposed:

(1) An initial estimate of parameters θ is given by $\tilde{\theta}$. It is obtained by the QMLE of

$$r_t = \mu + \rho r_{t-1} + \varepsilon_t \qquad \varepsilon_t = \sigma_t(\theta) z_t$$
$$\sigma_t^2(\theta) = \omega + \alpha_1 \varepsilon_{t-1}^2 \qquad \text{with} \qquad \theta = (\mu, \rho, \omega, \alpha_1)'$$

(2) The fitted residuals $\hat{\varepsilon}_t$ and the fitted variances $\hat{\sigma}_t^2(\theta)$ are used to compute the standardized residuals $\hat{z}_t(\theta) = \hat{\varepsilon}_t / \hat{\sigma}_t(\theta)$, with $E[\hat{z}_t(\theta)] = 0$ and $V[\hat{z}_t(\theta)] = 1$.

(3) The density $g(\hat{z}_t(\theta))$ is estimated using a non-parametric method. The estimated density is denoted $\hat{g}_t$.

(4) The log-likelihood is then computed as

$$\log(L_T(\theta; r_1, \ldots, r_T)) = \sum_{t=1}^{T} \log(\ell_t(\theta)) \quad \text{where} \quad \log(\ell_t(\theta)) = -\tfrac{1}{2}\log(\hat{\sigma}_t^2(\theta)) + \log(\hat{g}_t)$$

The log-likelihood is maximized, with $\hat{g}_t$ fixed, iterating (2)−(4) until convergence.

Semi-parametric estimation

The density is approximated using the histogram of z_t with knots $(n_1,...,n_{m-1})$ and heights $(p_1,..., p_{m-1})$. For a sample $(z_1,...,z_T)$ and an interval (a,b) divided in m sub-intervals of length q, the following optimization problem is solved

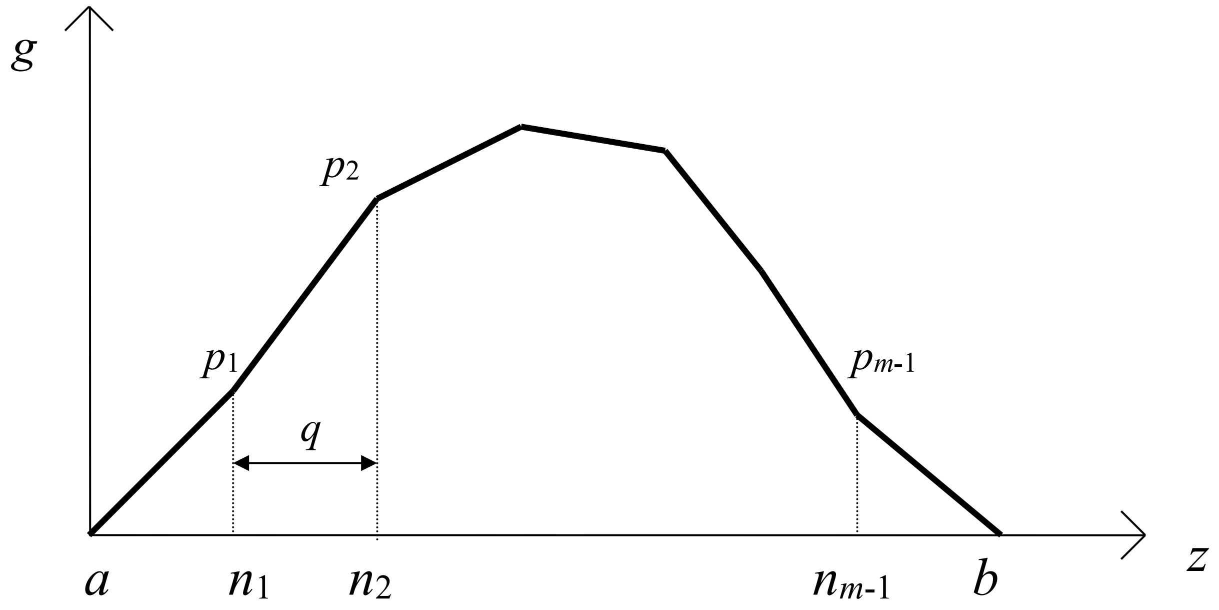
$$\max_{(p_1,...,p_{m-1})} \sum_{t=1}^{T} \log(g(z_t)) - \frac{\lambda}{q} \sum_{k=1}^{m-1} (p_{k+1} - 2p_k + p_{k-1})^2$$

subject to $q\sum_{k=1}^{m-1} p_k = 1$ and $p_k \geq 0$ for $k = 1,...,m-1$ $(p_0 = p_m = 0)$

and $$g(z) = \begin{cases} p_k + \dfrac{p_{k+1} - p_k}{q}(z - n_k) & \text{if } z \in [n_k; n_{k+1}) \\ 0 & \text{if } z \notin (n_0; n_m) \end{cases}$$

and λ is the penalty term to ensure smoothness of the estimate of $(p_1,..., p_{m-1})$.

Semi-parametric estimation



Semi-parametric method avoids some problems with distribution mis-specification, as using a non-normal distribution may lead to inconsistent parameter estimates if the distribution is incorrect.

At the same time, the semi-parametric estimator may be more efficient (i.e., with smaller standard error) than a fully non-parametric method.

Series expansion about the normal distribution

When the true pdf of a r.v. Z is unknown, yet believed to be close to a normal one, we can use an approximation of this pdf around the normal density

$$g(z\,|\,\eta) = \varphi(z)\, p_n(z\,|\,\eta)$$

where $\varphi(z)$ is the standard normal $N(0,1)$ and $p_n(z\,|\,\eta)$ is chosen so that $g(z\,|\,\eta)$ has the same first moments as the pdf of Z.

Special case of a series expansion: **Gram-Charlier type A expansion**.

Gallant and Tauchen (1989) use Gram-Charlier expansions to describe deviations from normality of innovations in a GARCH framework.

The approximating distribution is therefore

$$g(z\,|\,\eta) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \times \left[1 + \frac{m_3}{6}(z^3 - 3z) + \frac{m_4 - 3}{24}(z^4 - 6z^2 + 3)\right]$$

with $\eta = (m_3, m_4)'$ **(See Appendix for the derivation)**

Gram-Charlier distribution − Properties

GC expansions allow for additional flexibility over a normal distribution as they introduce skewness and kurtosis as unknown parameters.

The first four moments of Z are:

$$E[Z] = \int_{-\infty}^{\infty} z g(z \mid \eta) dz = 0 \qquad V[Z] = \int_{-\infty}^{\infty} z^2 g(z \mid \eta) dz = 1$$

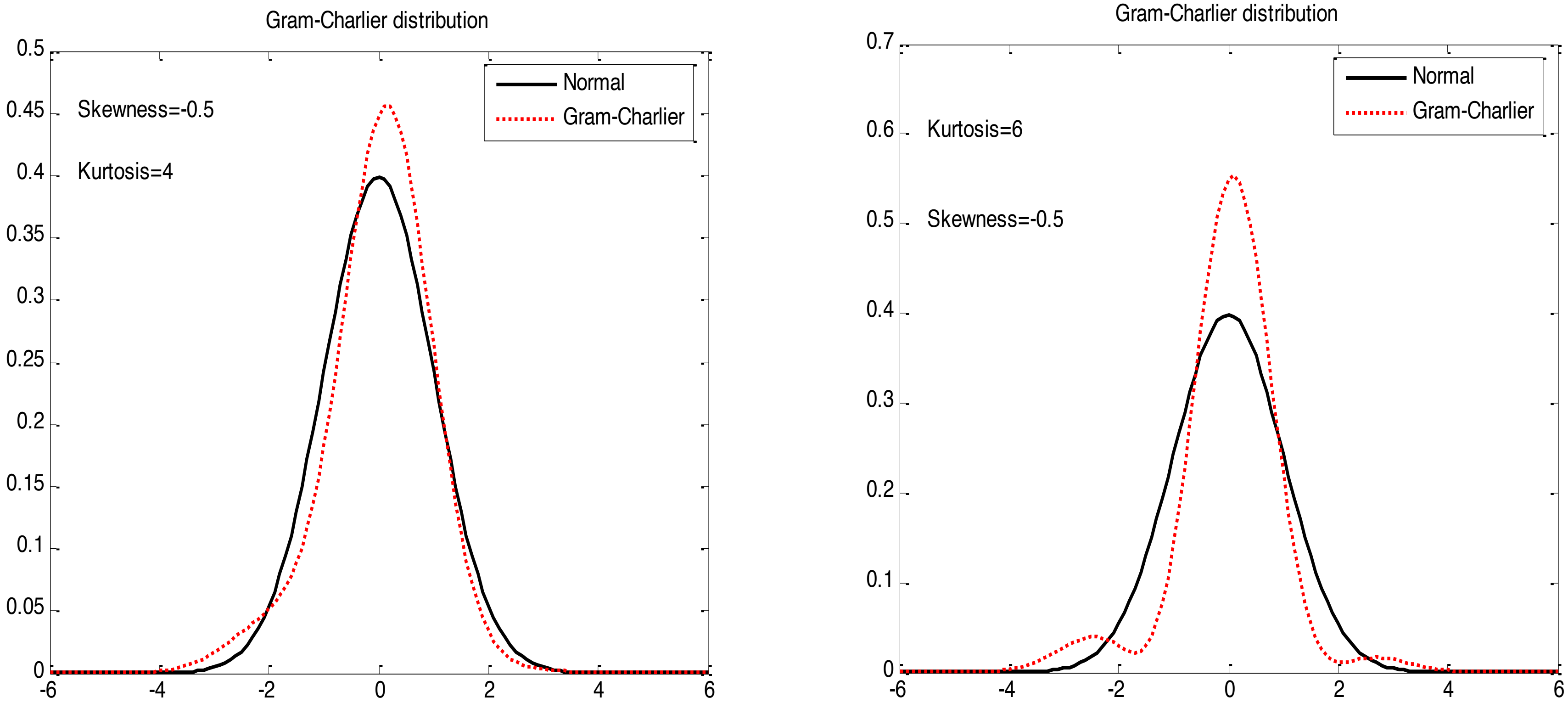
$$Sk[Z] = \int_{-\infty}^{\infty} z^3 g(z \mid \eta) dz = m_3 \qquad Ku[Z] = \int_{-\infty}^{\infty} z^4 g(z \mid \eta) dz = m_4$$

This result partly explains the success of Gram-Charlier expansions, as the two additional parameters m_3 and m_4 directly represent the third and fourth moments.

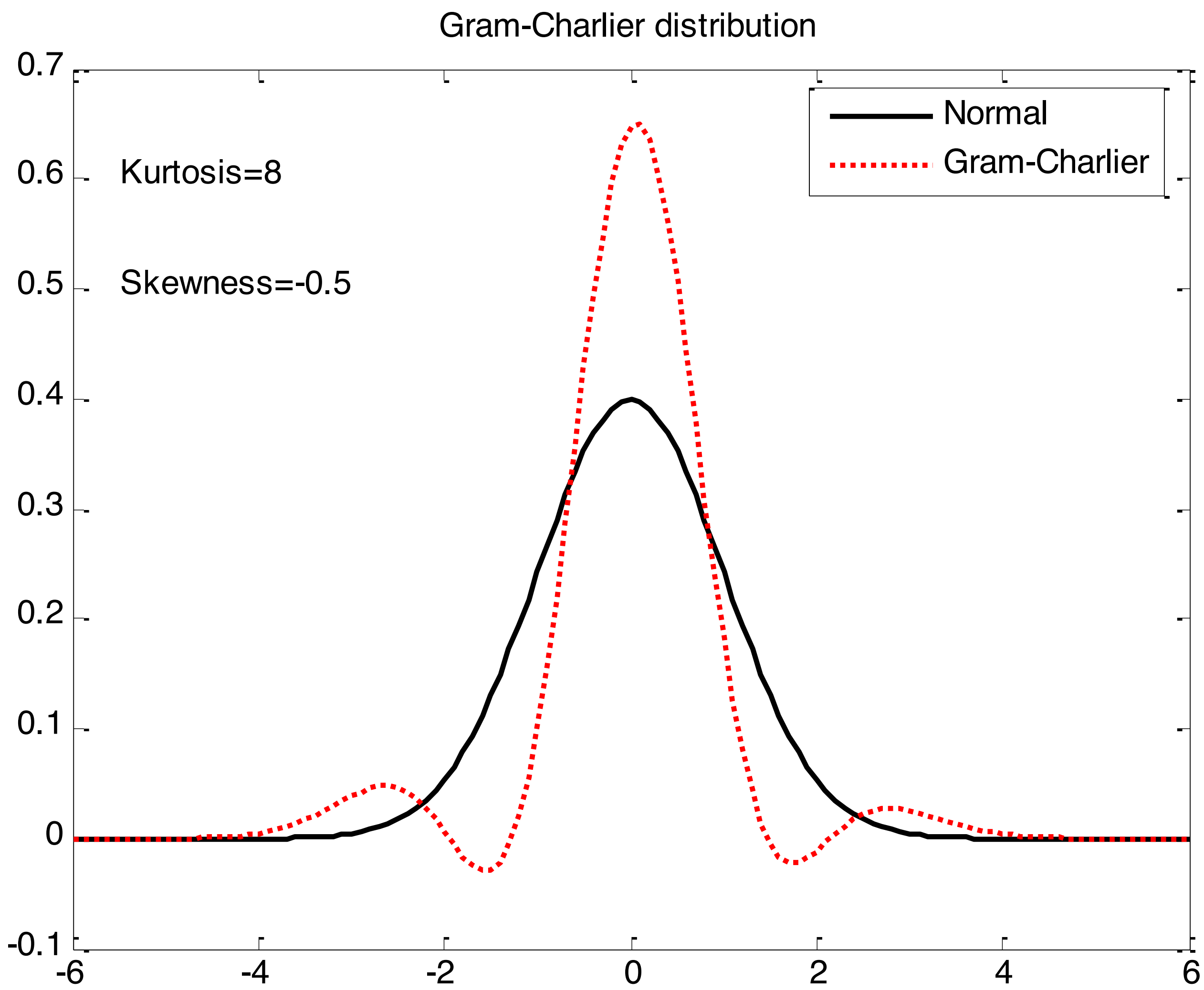
GC expansions have some drawbacks:

1. For other pairs, the pdf $g(z \mid \eta)$ may be multimodal.
2. For (m_3, m_4) far from normality (0,3), $g(z \mid \eta)$ may be negative for some z.
3. The domain of definition, for which the distribution is well defined, is small.

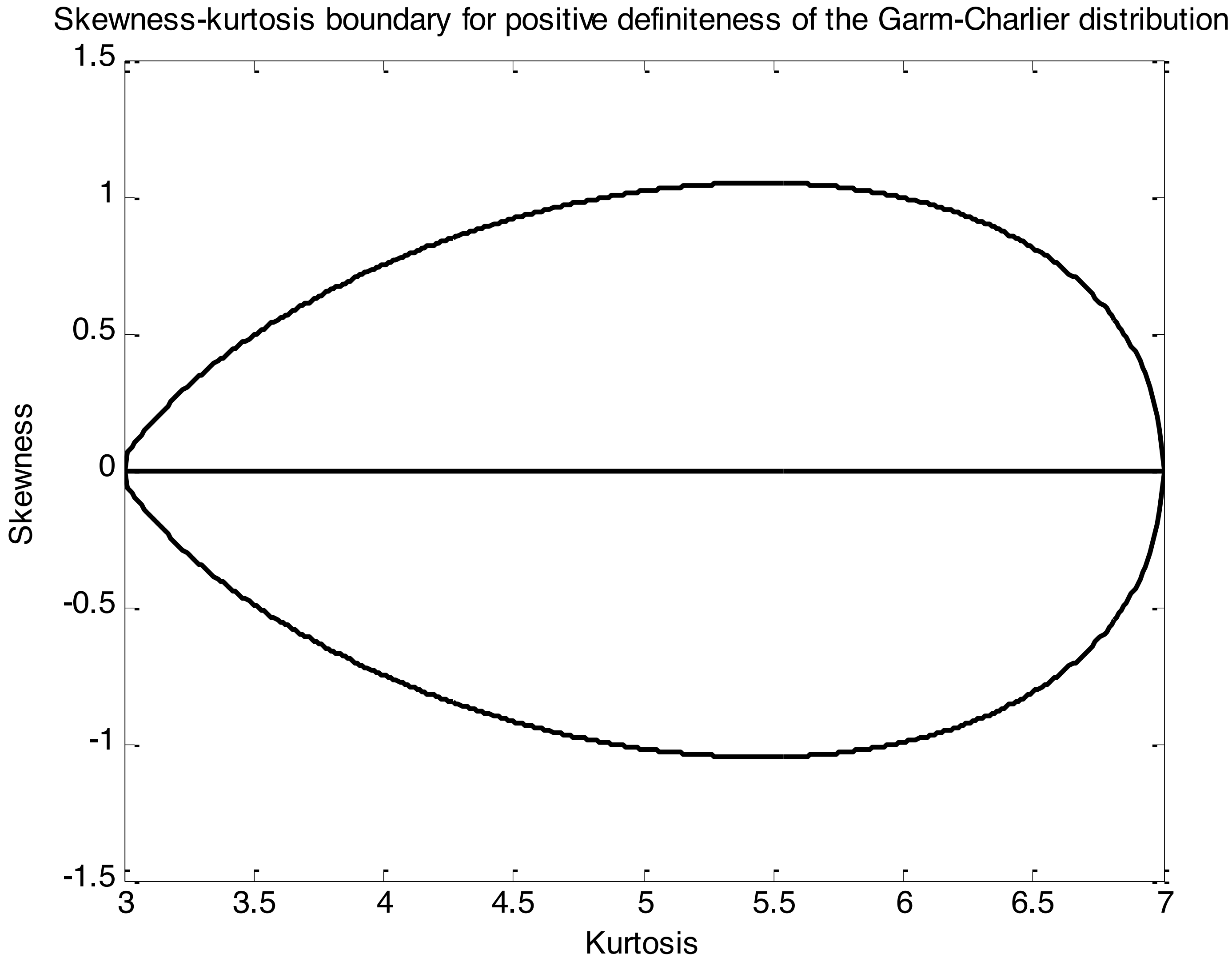
Gram-Charlier distribution − Illustration



Gram-Charlier distribution − Illustration



Gram-Charlier distribution − Domain of definition



Skewed Student *t* distribution

Bollerslev (1987, *REStat*): Student-*t* distribution captures the fat tails of financial returns. It is, however, a symmetric distribution. The *t* distribution is normalized

$$g(z_t \mid \eta) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} \qquad \text{where} \qquad \Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

so that $E[Z_t]=0$ and $V[Z_t]=1$.

Hansen (1994, *IER*): Skewed Student-*t* distribution captures the fat tails and the asymmetry of the distribution

He generalizes the Student-*t* distribution with asymmetry, while maintaining the assumption of a zero mean and unit variance.

Remark: By assuming that parameters depend on past realizations, he also shows that parameters, and subsequently higher moments, may be made time varying.

Skewed Student *t* distribution − Definition

Distribution for the standardized variable *Z*, with $E[Z_t]=0$ and $V[Z_t]=1$:

$$g(z_t \mid \eta) = b\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{\zeta_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}$$

where $$\zeta_t = \begin{cases}(bz_t+a)/(1-\lambda) & \text{if } z < -a/b \\ (bz_t+a)/(1+\lambda) & \text{if } z \geq -a/b\end{cases}$$

$$a = 4\lambda c\frac{\nu-2}{\nu-1} \qquad b^2 = 1+3\lambda^2-a^2 \qquad c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}$$

Shape parameters: $\eta=(\nu,\lambda)'$, with ν is the degree-of-freedom parameter $(2<\nu<\infty)$ and λ the asymmetry parameter $(-1<\lambda<1)$.

$\lambda=0$: traditional Student-*t* distribution / $\lambda=0$ and $\nu=\infty$: normal distribution.

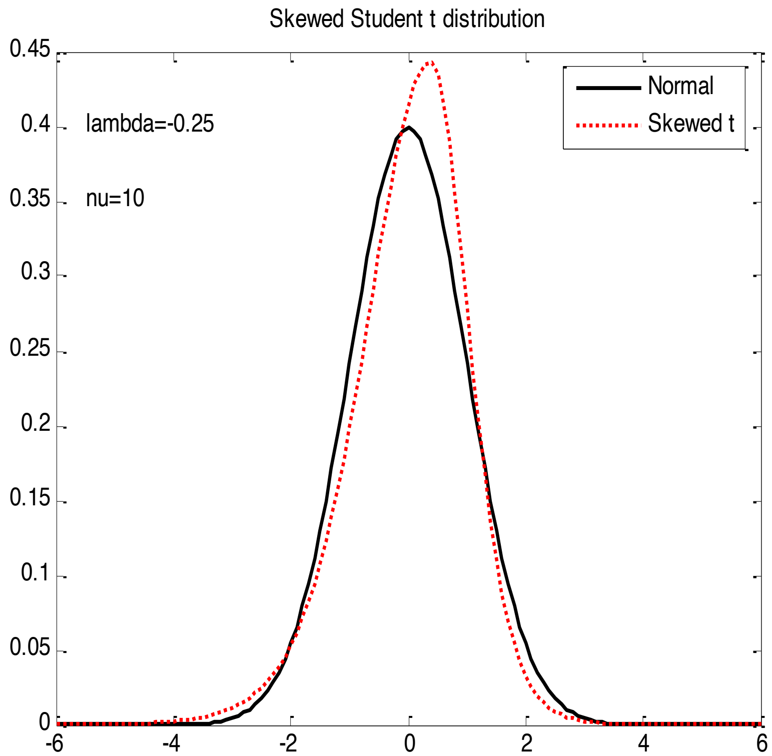
Skewed Student *t* distribution − Domain of definition

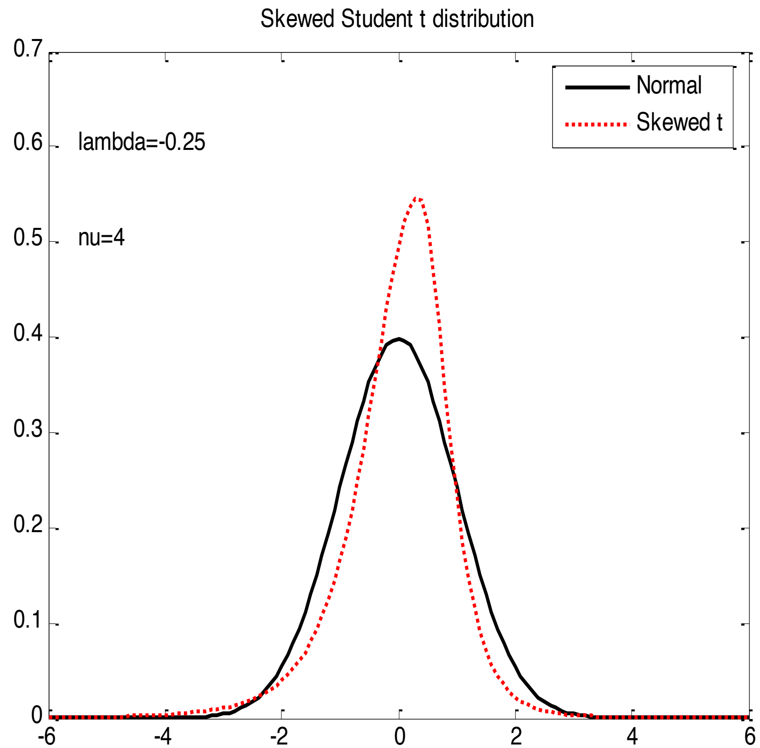
The density and the various moments do not exist for all parameters:

- The density is defined only for $\nu > 2$ and $-1 < \lambda < 1$.

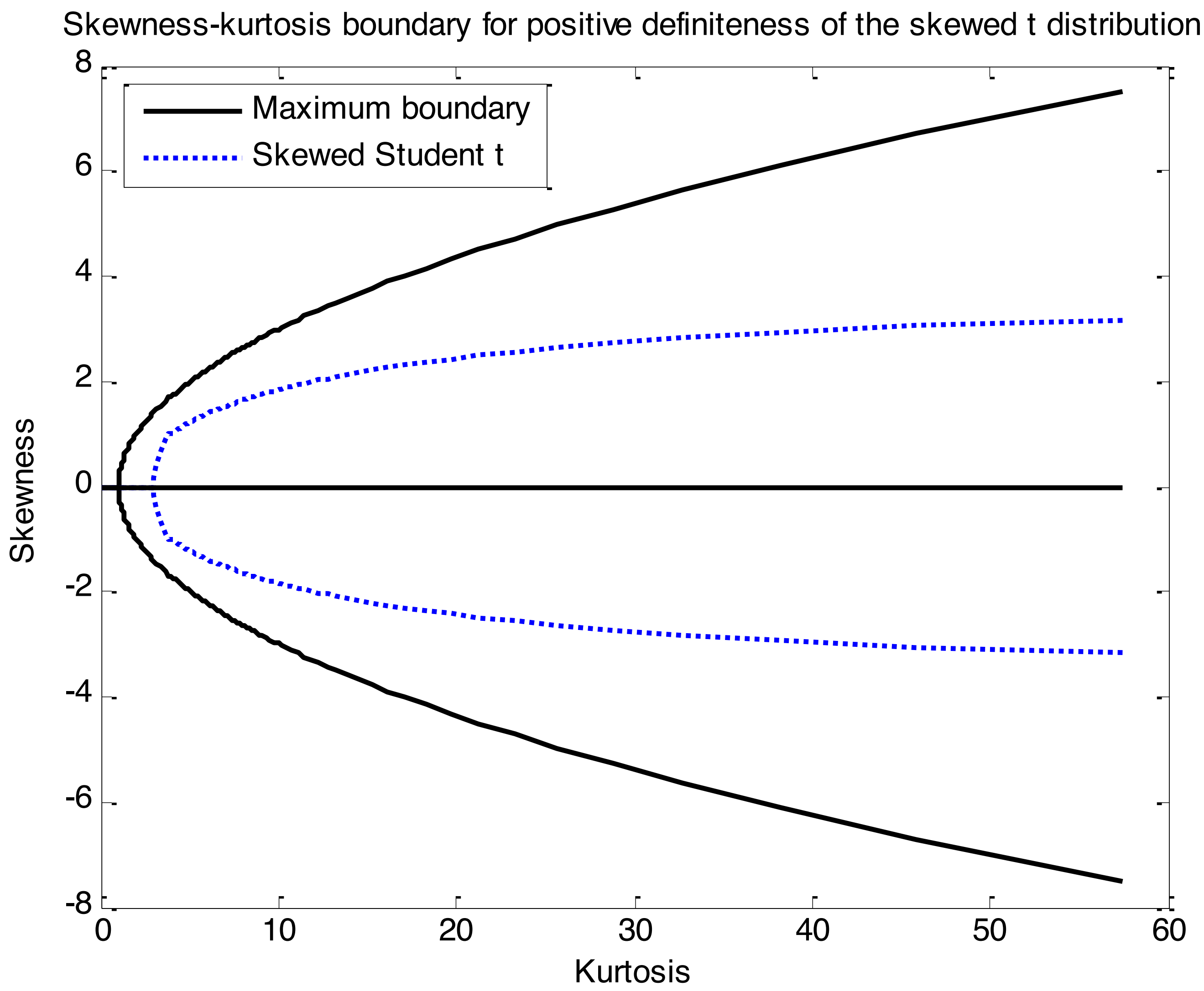
- Skewness exists if $\nu > 3$.

- Kurtosis exists if $\nu > 4$.





Skewed Student *t* distribution − Domain of definition



Objectives of the lecture

- Non-normal Distributions
- QMLE
- Existence of moments
- Modeling the conditional distribution

➔ **Adequacy tests**

Tests based on density forecasts – Proposition

Test of adequacy of the estimated distribution: based on the distance between the empirical distribution and the assumed (estimated) distribution. It is inspired by the Kolmogorov-Smirnov test used to test normality.

The empirical pdf is $f_t(z_t)$ and the sequence of density forecasts is $g_t(z_t)$. We define the probability integral transform as: $u_t = \int_{-\infty}^{z_t} g_t(y_t)dy_t = G_t(z_t)$.

Proposition (Diebold, Gunther, and Tay, 1998, *IER*). Suppose $\{z_t\}$ is generated from the distribution $\{f_t(z_t \mid I_{t-1})\}$ where $I_{t-1} = \{z_{t-1}, z_{t-2}, ...\}$. If a sequence of density forecasts $\{g_t(z_t \mid I_{t-1})\}$ coincides with $\{f_t(z_t \mid I_{t-1})\}$, then under some technical conditions, the sequence of probability integral transforms of $\{z_t\}$ with respect to $\{g_t(z_t \mid I_{t-1})\}$ is: $\{u_t\} \sim iid\ U(0,1)$.

In words: if the assumed distribution $g_t(z_t)$ is correct, then its probability integral transform should be $u_t \sim iid\ U(0,1)$.

Tests based on density forecasts – Sketch of the Proof

Since $u_t = G_t(z_t)$, one also has $z_t = G_t^{-1}(u_t)$, so that $f_t(z_t) = f_t(G_t^{-1}(u_t))$

Jacobian matrix: If we have $x = h(y)$, with density $f_X(x)$, then the density of Y is

$$f_Y(y) = f_X\big(h(y)\big) \times |J(x,y)| \qquad \text{with} \qquad J(x,y) = \frac{\partial h}{\partial y}$$

Here, we have $f_X(x) = f_t(z_t)$ and $h(y) = G_t^{-1}(u_t)$, so that

$$\frac{\partial h}{\partial y} = \frac{\partial G_t^{-1}(u_t)}{\partial u_t} = \frac{1}{g_t(G_t^{-1}(u_t))}$$

Therefore, the density of u_t is:

$$q_t(u_t) = f_t(z_t) \times |J(z_t, u_t)| = f_t(G_t^{-1}(u_t)) \times \left|\frac{\partial G_t^{-1}(u_t)}{\partial u_t}\right| = \frac{f_t(G_t^{-1}(u_t))}{g_t(G_t^{-1}(u_t))}$$

If $f_t(z_t) = g_t(z_t)$, then: $q_t(u_t) = U(0,1)$.

Tests based on density forecasts – Test Procedure

The adequacy test is based on two steps:

1) Test whether u_t is serially correlated, using a standard LM test. For this purpose, we regress $(u_t - \bar{u})^i$ on K lags of the variable.

$$(u_t - \bar{u})^i = a_0 + a_1 (u_{t-1} - \bar{u})^i + \ldots + a_K (u_{t-K} - \bar{u})^i + e_t \qquad \text{, for } i = 1,\ldots,4$$

The LM test statistics, $T \times R^2$, is asymptotically distributed, under the null, as a $\chi^2(K)$.

2) Test the null that u_t is $U(0,1)$. We cut the empirical and theoretical distributions into N cells and test if the two distributions are the same, using Pearson's test statistic:

$$DGT(N) = \sum_{n=1}^{N} \frac{(F_n - T/N)^2}{T/N}$$

where F_n is the number of observations in cell n and T/N is the expected number of observations under the null. Under the null, we have $DGT(N) \overset{a}{\sim} \chi^2(N-1)$.

Tests based on density forecasts – Confidence Bands

A simple estimate of the standard error of the estimated number of observations is based on the **binomial distribution**.

Under the null, the statistics F_n can be viewed as drawn from a binomial distribution $B(T, p)$, where T is the number of realizations and $p = 1/N$ is the probability to fall in cell n. We have

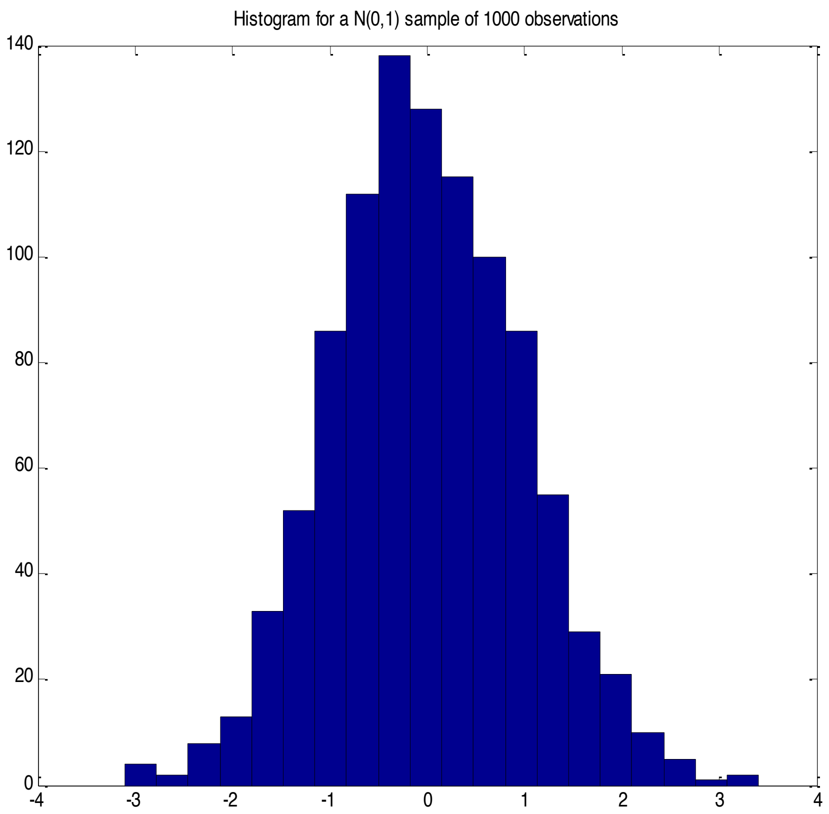
$$E[F_n] = Tp = \frac{T}{N} \qquad \text{and} \qquad V[F_n] = Tp(1-p) = T\frac{1}{N}\left(1-\frac{1}{N}\right).$$

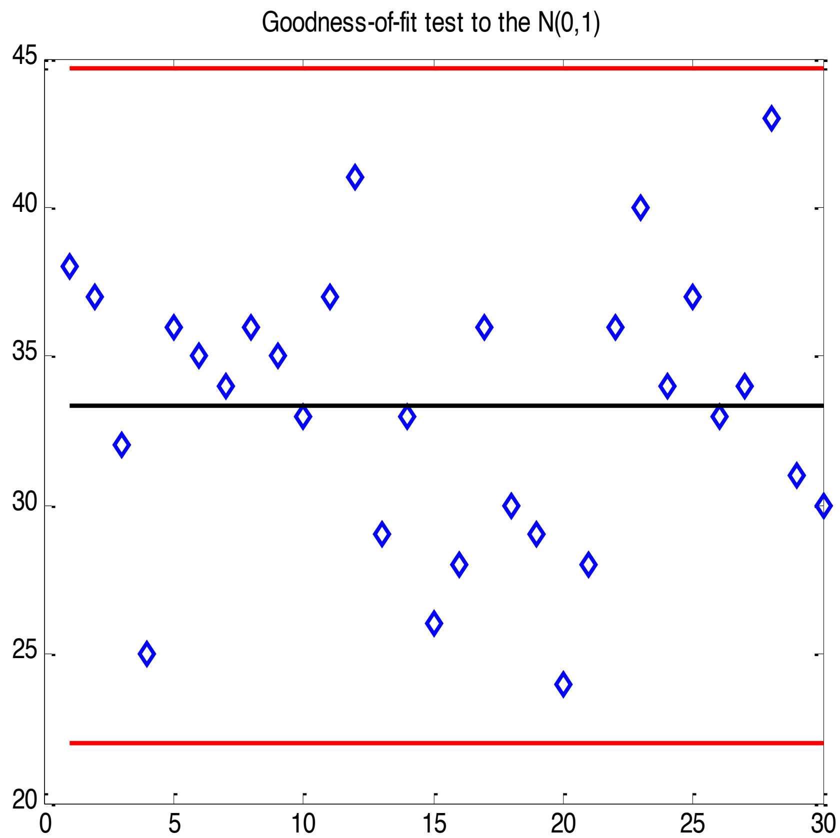
The confidence bands in the next slides are constructed using this expression with

$$\left[E[F_n] \pm 2\sqrt{V[F_n]}\right] = \left[\frac{T}{N} \pm 2\sqrt{T\frac{1}{N}\left(1-\frac{1}{N}\right)}\right]$$

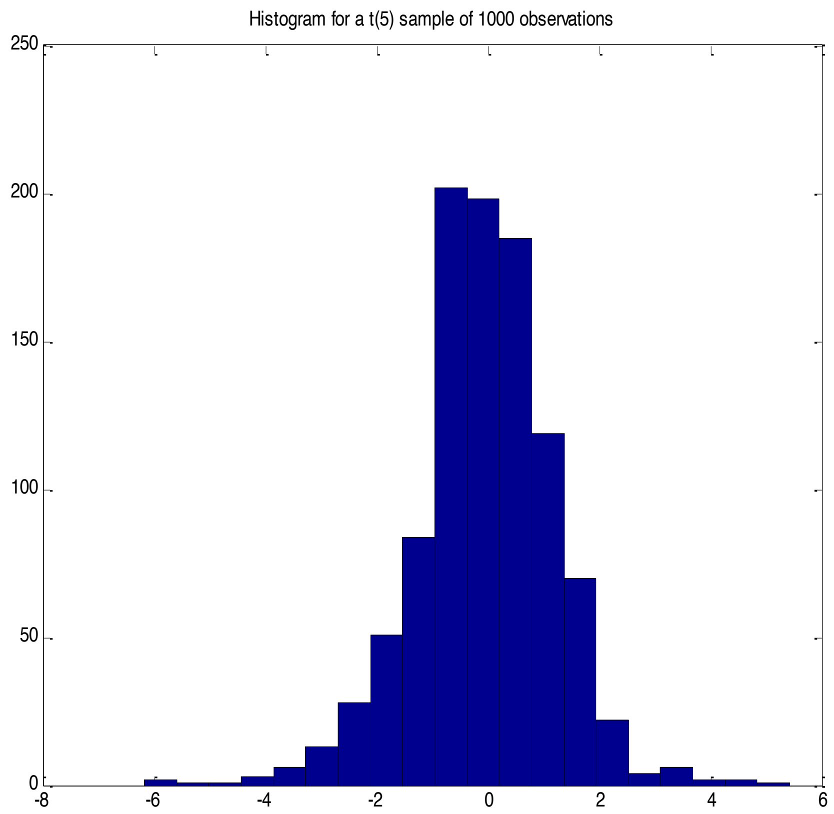
Remark: In Pearson's test statistic above, the Binomial distribution $B(T, p)$ is approximated by a Normal $N(Tp, Tp)$.

Example

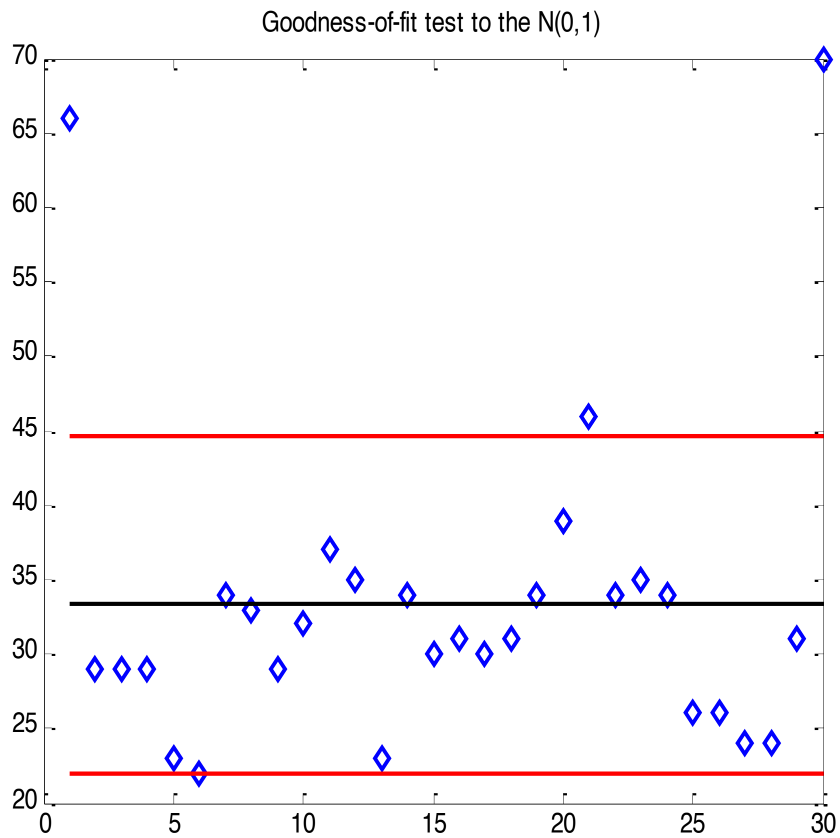

Histogram for a N(0,1) sample of 1000 observations
140
120
100
80
60
40
20
0
-4
-3
-2
-1
0
1
2
3
4


Goodness-of-fit test to the N(0,1)
45
40
35
30
25
20
0
5
10
15
20
25
30

Example

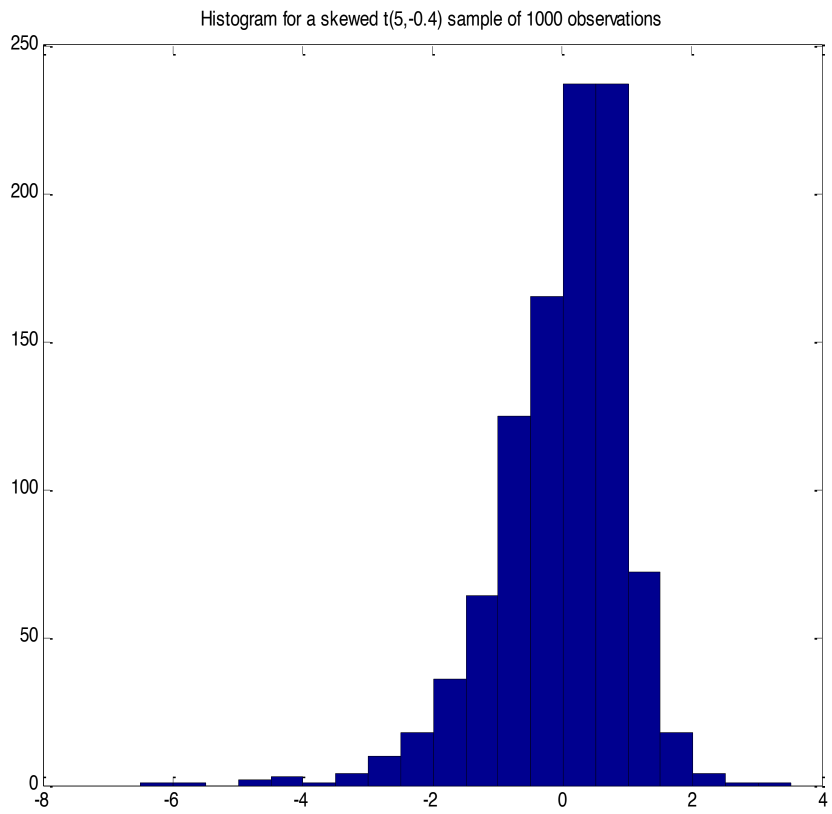


Histogram for a t(5) sample of 1000 observations
250
200
150
100
50
0
-8
-6
-4
-2
0
2
4
6

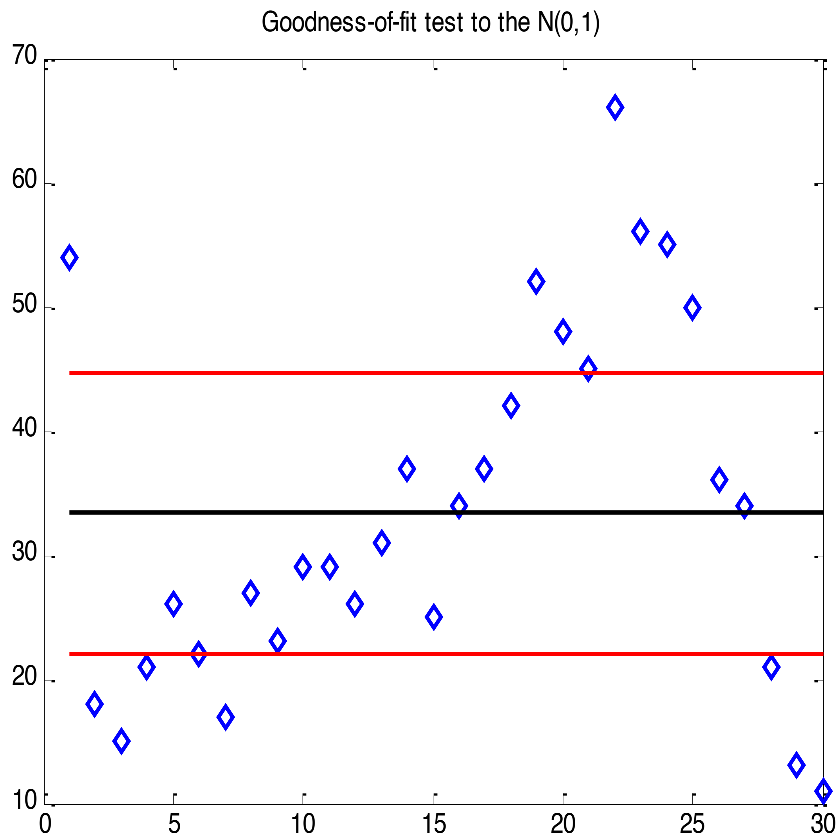


Goodness-of-fit test to the N(0,1)
70
65
60
55
50
45
40
35
30
25
20
0
5
10
15
20
25
30

Example



Histogram for a skewed t(5,-0.4) sample of 1000 observations
250
200
150
100
50
0
-8
-6
-4
-2
0
2
4



Goodness-of-fit test to the N(0,1)
70
60
50
40
30
20
10
0
5
10
15
20
25
30

Example

$DGT(30)$ statistics for the null hypothesis that the actual distribution is a $N(0,1)$.
The simulation is based on $T = 1000$ observations of the assumed distribution.

Simulated distribution	**$N(0,1)$**	**$t(5)$**	**Skewed $t(5,-0.4)$**
Statistics	22.84	99.74	151.04
p-value	0.7793	0	0

Appendix: Derivation of the Gram-Charlier distribution

When the true pdf of a r.v. Z is unknown, yet believed to be close to a normal one, we can use an approximation of this pdf around the normal density

$$g(z \mid \eta) = \varphi(z) p_n(z \mid \eta)$$

where $\varphi(z)$ is the standard normal density with zero mean and unit variance and where $p_n(z \mid \eta)$ is chosen so that $g(z \mid \eta)$ has the same first moments as the pdf of z.

Gallant and Tauchen (1989) use a **Gram-Charlier type A expansion** to describe deviations from normality of innovations in a GARCH framework. The polynomial writes:

$$g(z \mid \eta) = \varphi(z) + c_1 \frac{\partial \varphi(z)}{\partial z} + c_2 \frac{\partial^2 \varphi(z)}{\partial z^2} + \cdots + c_i \frac{\partial^i \varphi(z)}{\partial z^i} + \cdots = \sum_{i=1}^{\infty} c_i H_i(z) \varphi(z)$$

where $H_i(z) = (-1)^i \dfrac{\partial^i \varphi(z)}{\partial z^i} \dfrac{1}{\varphi(z)}$ is called Hermite polynomial of order i.

Derivation of the Gram-Charlier distribution

In particular,

$H_0 = 1 \qquad H_1 = z \qquad H_2 = z^2 - 1 \qquad H_3 = z^3 - 3z \qquad H_4 = z^4 - 6z^2 + 3$

We can verify that Hermite polynomials have an orthogonal property:

$$\int_{-\infty}^{\infty} H_i(z) H_j(z) \varphi(z)\, dz = \begin{cases} i! & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

We use this property to compute the constant c_i by integrating

$$\int_{-\infty}^{\infty} H_j(z)\varphi(z)\, dz = \int_{-\infty}^{\infty} H_j(z) \sum_{i=1}^{\infty} c_i H_i(z)\varphi(z)\, dz = c_j \int_{-\infty}^{\infty} H_j(z) H_j \varphi(z)\, dz = j! \times c_j$$

so that $c_j = \dfrac{1}{j!}\int_{-\infty}^{\infty} H_j(z)\varphi(z)\, dz$ with

$c_0 = 1 \qquad c_1 = 0 \qquad c_2 = (m_2 - 1)/2 \qquad c_3 = m_3 / 3! \qquad c_4 = (m_4 - 6m_2 + 3)/4!$

Derivation of the Gram-Charlier distribution

Finally, we obtain

$$g(z \mid \eta) \approx \varphi(z)\left[1+\frac{1}{2}(m_2-1)H_2(z)+\frac{1}{6}m_3H_3(z)+\frac{1}{24}(m_4-6m_2+3)H_4(z)\right]$$

Since Z has been standardized, we have $m_2 = 1$ so that

$$g(z \mid \eta) \approx \varphi(z)\left[1+\frac{m_3}{6}H_3(z)+\frac{m_4-3}{24}H_4(z)\right] \qquad \text{with } \eta = (m_3, m_4)'$$

GC expansions allow for additional flexibility over a normal distribution since they introduce the skewness and kurtosis of the distribution as unknown parameters.

The approximating distribution is therefore

$$g(z \mid \eta) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}z^2\right)\times\left[1+\frac{m_3}{6}(z^3-3z)+\frac{m_4-3}{24}(z^4-6z^2+3)\right]$$