Financial Data Analysis

Markus Haas

Summer term 2014

July 15, 2014
Financial Data Analysis

- E-mail: markus.haas@qber.uni-kiel.de

- Organizational/exam issues: cornelia.hupfer@vwl.uni-freiburg.de (Cornelia Hupfer)
Course Outline

• Introduction: Asset return definitions

• Statistical properties of financial returns
  – Marginal distribution: **fat tails** and **skewness**
  – Alternative distributions for returns
  – Temporal properties of returns (volatility clustering)
  – Time–varying dependence structure (correlations)

• Introduction to time series analysis

• Statistical models for returns:
  (i) Univariate GARCH models and extensions
  (ii) Multivariate GARCH models
  (iii) Realized volatility

• Further topics
Textbooks


Other (potentially) useful books

- Andersen/Davis/Kreiß/Mikosch, eds. (2009), Handbook of Financial Time Series, Springer
- Bauwens/Hafner/Rombouts, eds. (2012), Handbook of Volatility Models and Their Applications, Wiley
- Enders (2010), Applied Econometric Time Series, 3e, Wiley
- Paolella (2006), Fundamental Probability, Wiley
Asset Return Definitions

• Let \( P_t \) be the price of an asset at time \( t \) (stock, stock index, exchange rate,...).

• If there are no dividends or other payments, the (single–period) **discrete net return** is

\[
R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1,
\]

whereas the discrete **gross return** is

\[
1 + R_t = \frac{P_t}{P_{t-1}}.
\]

• Often the returns defined in (1) and (2) are multiplied by 100 to be interpretable in terms of **percentage returns**.
Asset Return Definitions

- If an asset pays a dividend, $D_t$, between time $t-1$ and $t$, then the return is

$$R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t - P_{t-1}}{P_{t-1}} + \frac{D_t}{P_{t-1}},$$  \hspace{1cm} (3)$$

where the term $(P_t - P_{t-1})/P_{t-1}$ is the capital gain, and $D_t/P_{t-1}$ is the dividend yield.
Asset Return Definitions

- The continuously compounded or log–return of an asset is

\[ r_t = \log \left( \frac{P_t}{P_{t-1}} \right) = \log P_t - \log P_{t-1} = \log(1 + R_t), \quad (4) \]

and the percentage log–return is (4) multiplied by 100.

- Its name derives from the fact that (4) can be rewritten as

\[ P_t = P_{t-1} e^{r_t}, \quad (5) \]

i.e., \( r_t \) is the rate of return under continuous compounding between time \( t - 1 \) and time \( t \), i.e.,

\[ P_{t-1} \left[ \lim_{n \to \infty} \left(1 + \frac{r_t}{n}\right)^n \right] = P_{t-1} e^{r_t} = P_t. \quad (6) \]
Asset Return Definitions

• Since $e^x \geq 1 + x$, we always have

$$r_t = \log(1 + R_t) \leq R_t.$$  \hspace{1cm} (7)

• For small returns,\(^1\)

$$r_t = \log(1 + R_t) = R_t - \frac{R_t^2}{2} + \frac{R_t^3}{3} - \frac{R_t^4}{4} + \cdots$$  \hspace{1cm} (8)

$$\approx R_t,$$  \hspace{1cm} (9)

so that $r_t$ may serve as a reasonable approximation to the discrete return $R_t$.

\(^1\)Note that expansion (8) is valid only for $R_t \in (-1, 1]$; the approximation is very good at least for $R_t$ between $-0.1$ (−10%) and 0.1 (10%), cf. Table 1.
### Asset Return Definitions

- Examples of daily or weekly returns are very rarely outside the range from $-10\%$ to $10\%$, and then the choice of the return definition will be of minor importance.

#### Table 1: Discrete and continuous returns

<table>
<thead>
<tr>
<th></th>
<th>100 \times R_t</th>
<th>100 \times \log(1 + R_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1.00$</td>
<td>$-1.00$</td>
<td>$-1.01$</td>
</tr>
<tr>
<td>$-2.50$</td>
<td>$-2.50$</td>
<td>$-2.53$</td>
</tr>
<tr>
<td>$-5.00$</td>
<td>$-5.00$</td>
<td>$-5.13$</td>
</tr>
<tr>
<td>$-10.00$</td>
<td>$-10.00$</td>
<td>$-10.54$</td>
</tr>
<tr>
<td>$-15.00$</td>
<td>$-15.00$</td>
<td>$-16.25$</td>
</tr>
<tr>
<td>$1.00$</td>
<td>$1.00$</td>
<td>$1.00$</td>
</tr>
<tr>
<td>$2.50$</td>
<td>$2.50$</td>
<td>$2.47$</td>
</tr>
<tr>
<td>$5.00$</td>
<td>$5.00$</td>
<td>$4.88$</td>
</tr>
<tr>
<td>$10.00$</td>
<td>$10.00$</td>
<td>$9.53$</td>
</tr>
<tr>
<td>$15.00$</td>
<td>$15.00$</td>
<td>$13.98$</td>
</tr>
</tbody>
</table>

Continuous *percentage* return $r_t = 100 \times \log(1 + R_t)$. 
discrete return $R_t$

log-return $\log(1 + R_t)$
Asset Return Definitions

- Both return definitions have their advantages and disadvantages.

- E.g., in time series analysis, continuously compounded returns are often used because multi–period returns are sums of single–period returns.

- That is, if $r_{t:t+\tau}$ denotes the (multi–period) return from time $t$ to time $t + \tau$, we have

\[
r_{t:t+\tau} = \log \left( \frac{P_{t+\tau}}{P_t} \right)
= \log \left( \frac{P_{t+\tau}}{P_{t+\tau-1}} \frac{P_{t+\tau-1}}{P_{t+\tau-2}} \cdots \frac{P_{t+2}}{P_{t+1}} \frac{P_{t+1}}{P_t} \right)
= \log \left( \frac{P_{t+1}}{P_t} \right) + \log \left( \frac{P_{t+2}}{P_{t+1}} \right) + \cdots + \log \left( \frac{P_{t+\tau}}{P_{t+\tau-1}} \right)
= \sum_{i=1}^{\tau} r_{t+i}.
\]
Asset Return Definitions

- This is not the case for the discrete return, where

\[ R_{t:t+\tau} = \prod_{i=1}^{\tau} (1 + R_{t+i}) - 1, \]

and sums of random variables can be more easily handled than products.

- Also, due to limited liability, most assets have a lower bound of zero, i.e., for the discrete return,

\[ R_t \geq -1, \]

which is easier to preserve in models based on continuous returns.\(^2\)

\(^2\)However, distributions with unbounded support, such as the normal, may still be used as an approximation for discrete returns, since, when fitted to return data, the implied probability of a loss larger than 100% will be negligible. The same reasoning of course applies to basically any application of such distributions to real world data, which are usually restricted to a finite-length interval.
Asset Return Definitions

• Discrete returns have a very convenient property in portfolio analysis not shared by continuous returns:

• Let
  - $P_{it}$ be the price of asset $i$ at time $t$, and
  - $R_{it}$ be the return of asset $i$,

  $i = 1, \ldots, N$, where $N$ is the number of assets in the portfolio.

• At time $t - 1$, we have $n_i$ units of asset $i$ in the portfolio.
Asset Return Definitions

- The value of the portfolio at time $t - 1$, $P_{p,t-1}$, is thus

$$P_{p,t-1} = \sum_{i=1}^{N} n_i P_{i,t-1},$$

(11)

and asset $i$’s portfolio weight (i.e., the fraction of wealth invested in asset $i$) is

$$x_i = \frac{n_i P_{i,t-1}}{\sum_{j=1}^{N} n_j P_{j,t-1}}, \quad i = 1, \ldots, N,$$

(12)

with

$$\sum_{i=1}^{N} x_i = 1.$$ 

(13)
Asset Return Definitions

- The portfolio return $R_{p,t}$ is given by

$$R_{t,p} = \frac{P_{p,t} - P_{p,t-1}}{P_{p,t-1}}$$

$$= \frac{\sum_i n_i P_{it} - \sum_i n_i P_{i,t-1}}{\sum_i n_i P_{i,t-1}}$$

$$= \frac{\sum_i n_i (P_{it} - P_{i,t-1})}{\sum_i n_i P_{i,t-1}}$$

$$= \sum_{i=1}^N \frac{n_i P_{i,t-1}}{\sum_i n_i P_{i,t-1}} \frac{P_{it} - P_{i,t-1}}{P_{i,t-1}}$$

$$= \sum_{i=1}^N x_i R_{i,t}.$$ 

- That is, the portfolio return is a linear combination of the components' returns.
Asset Return Definitions

- For the continuous return, in general,

\[
r_{p,t} = \log \left( \frac{P_{p,t}}{P_{p,t-1}} \right) = \log(1 + R_{p,t})
\]

\[
= \log \left( \sum_i x_i (1 + R_{it}) \right)
\]

\[
\neq \sum_i x_i \log(1 + R_{it})
\]

\[
= \sum_{i=1}^{N} x_i r_{it},
\]

i.e., the linear combination of continuously compounded asset returns is not the continuously compounded portfolio return.
Asset Return Definitions

• However, for small returns, the difference is again moderate, and the approximation

\[ r_{p,t} \approx \sum_{i=1}^{N} x_i r_{it} \]  \hspace{1cm} (14)

is also frequently used.
Basic Statistical Properties of Returns: Marginal (Unconditional) Return Distribution

- Due to the uncertain nature of the returns of speculative assets, they are best treated as random variables.

- The traditional assumption that has long dominated empirical finance was that log–returns over “longer” time intervals (i.e., daily or longer) are approximately normally distributed.

- The rationale behind this view is as follows:

- Daily log–returns, for example, are the sum of a large number of intraday returns.

- Appealing to the central limit theorem, Osborne (1959) argued in a classical article that

\[ \underbrace{\text{under fairly general conditions \ldots}}_{\text{Brownian Motion in the Stock Market, Operations Research 7, 145-173.}} \text{ we can expect that the distribution function of } r_t \text{ will be normal.} \]
Marginal (Unconditional) Return Distribution

• Is the normal assumption reasonable?

• To check this, we have to compare the normal with the empirical return distribution.

• *Kernel density estimates* and QQ (quantile–quantile) plots are frequently employed as a quick visual check.
Kernel Density Estimation

• The kernel density estimator is a nonparametric estimator of the density and can be viewed as a “smoothed” histogram.

• It looks more like a probability density function than a histogram and does not depend on the number and location of bins.

• This smoothed histogram can then be used to quickly (and informally) compare the empirical distribution with a fitted normal distribution (or any other hypothesized distribution).
Kernel Density Estimation

- Want to estimate the density of random variable $X$.

- The density of $X$, $f(x)$, is the derivative of the cdf, $F(x)$,

\[ F(x) = \Pr(X \leq x), \tag{15} \]

that is,

\[ f(x) = \lim_{h \to 0} \frac{1}{2h} \Pr(x - h < X < x + h). \tag{16} \]

- Thus we may attempt to estimate the density by choosing a small $h$ and setting

\[ \hat{f}(x) = \frac{1}{2hT} \left\{ \text{number of observations in } (x - h, x + h) \right\}, \tag{17} \]

where $T$ is the size of our sample.
Kernel Density Estimation

• For a sample $X_1, X_2, \ldots, X_T$, estimator (17) can be written as

$$\hat{f}(x) = \frac{1}{T} \sum_{t=1}^{T} \frac{1}{h} w \left( \frac{x - X_t}{h} \right),$$  \hspace{1cm} (18)$$

where the weight function

$$w(x) = \begin{cases} 
\frac{1}{2} & \text{if } |x| < 1 \\
0 & \text{otherwise}
\end{cases} \hspace{1cm} (19)$$

• That is, the estimator is obtained by placing a box of width $2h$ and height $\frac{1}{2Th}$ on each observation and then adding them up.

• How does this look like?
- This has been produced using a (pseudo) random sample of size $T = 100$ from a $N(0,1)$ distribution and $h = \frac{1}{2}$. 

![Graph of N(0,1) density and estimate with $h = 0.5$.]
Kernel Density Estimation

- The density estimator (17) has jumps at points $X_t \pm h$, $t = 1, \ldots, T$, and zero derivative everywhere else, leading to a ragged picture which may be undesirable.

- An intuitive motivation for the kernel estimator is thus to modify the estimator above by replacing the weight function (19) with a “nicer” kernel function $K(x)$ which leads to a smooth estimator.

- E.g., a Gaussian kernel is often used, such that

$$K(x) = \phi(x) = (2\pi)^{-1/2} \exp\{-x^2/2\}. \quad (20)$$

- In general, the kernel must be a density, i.e., $K(x) \geq 0$ and $\int K(x)dx = 1$. 
Kernel Density Estimation

- The kernel density estimator at a point \( x \) is

\[
\hat{f}(x) = \frac{1}{Th} \sum_{t=1}^{T} K \left( \frac{x - X_t}{h} \right),
\]

where \( K \) is the kernel function.

- Parameter \( h \) is the bandwidth or smoothing parameter which determines the smoothness of the estimator.

- A larger \( h \) produces a smoother density estimate.

- For example, consider daily log–returns of the DAX 30 index from January 2000 to May 2013 (\( T = 3477 \) observations), with \( h = 0.05 \) and \( h = 0.5 \).
density of daily DAX returns

kernel (h = 0.05)
Kernel Density Estimation

- How to choose $h$?

- A small $h$ allows the density estimator to capture the fine structure of the true density, but also introduces random variation.

- On the other hand, a large $h$ reduces random variation but may hide important details of the density.

- Thus, as is often the case in statistics, there is a bias/variance trade-off.
Kernel Density Estimation

• Trying to choose $h$ optimally (in the MSE\(^4\)) sense shows that the optimal $h$ depends on $T$ (the sample size), the kernel, and, unfortunately but not surprisingly, the true density (which is unknown).

• However, it turns out\(^5\) that, with a Gaussian kernel,\(^6\) choosing

$$h = 0.9AT^{-1/5}$$ \hspace{1cm} (22)

gives good results for a wide range of densities, where

$$A = \min\{\text{sample standard deviation, interquartile range}/1.34\},$$ \hspace{1cm} (23)

and the interquartile range is the difference between the 0.75–quantile and the 0.25–quantile.

\(^4\)Mean squared error.
\(^5\)See the discussion in Silverman, 1986, Chapter 3.
\(^6\)And a random (iid) sample, which will not hold for most financial data.
Kernel Density Estimation

- In financial applications, the sample size is often large enough so that the estimator is not too sensitive to changes of $h$ within reasonable limits.

S&P 500 index level (daily), January 2000 to October 2011
S&P 500 index returns (daily), January 2000 to October 2011
Kernel Density Estimation

- We compare the kernel estimate with a fitted normal, i.e., with the density

\[ f(x) = \frac{1}{\sqrt{2\pi \hat{\sigma}}} \exp \left\{ -\frac{(x - \hat{\mu})^2}{2\hat{\sigma}^2} \right\}, \]

where

\[ \hat{\mu} = \bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t \]  \hspace{1cm} (24)

and

\[ \hat{\sigma}^2 = s^2 = \frac{1}{T - 1} \sum_{t=1}^{T} (r_t - \bar{r})^2 \] \hspace{1cm} (25)

are the sample mean and variance, respectively.
Density of S&P 500 returns

- **empirical (kernel)**
- **fitted normal**
The log–density helps to better detect differences in the tails:
DAX 30 index returns (daily), January 2000 to October 2011
Density of DAX 30 returns

- empirical (kernel)
- fitted normal
Basic Statistical Properties of Returns: Excess Kurtosis (Thick Tails)

- Financial Returns at higher frequencies (higher than a month at least) are not normally distributed.

- In particular, they have much more probability mass in the center and the tails (fat tails) than a normal distribution with the same variance.

- The empirical density exhibits excess kurtosis relative to the normal distribution.

- This implies, among other things, that the probability of large losses is much higher than under the Gaussian assumption, which is of considerable interest for many financial applications.
Basic Statistical Properties of Returns: Thick Tails

- A further simple (but useful) tool for detecting departures from normality (or any other hypothesized distribution) are QQ (quantile–quantile) plots.

- This is a scatter plot of the empirical quantiles against the theoretical quantiles implied by a hypothesized distribution (e.g., the normal distribution).

- If the theoretical quantiles are from the normal distribution, the QQ plot is also referred to as normal probability plot.
Sample cdf, sample quantiles, QQ plots

- Given a sample of length \( T, X_1, X_2, \ldots, X_T \).

- The \textit{sample} or \textit{empirical cdf} \( F_T(x) \) is defined to be the proportion of values in the sample that are not larger than \( x \), i.e.,

\[
F_T(x) = \frac{1}{T} \sum_{t=1}^{T} I(X_t \leq x) = \frac{\text{number of } X_t \text{ not larger than } x}{T},
\]

where \( I \) is the indicator function such that

\[
I(X_t \leq x) = \begin{cases} 
1 & \text{if } X_t \leq x \\
0 & \text{otherwise}
\end{cases}
\]
sample cdf of the DAX returns
cdf of the DAX returns

- Sample CDF
- Fitted Normal
cdf of the DAX returns (left tail)

- Sample CDF
- Fitted Normal
Sample quantiles

- Recall that the $\alpha$–quantile of a random variable $X$, denoted by $\xi_\alpha$, is the smallest number $\xi$ such that

\[ F(\xi) \geq \alpha, \quad 0 < \alpha < 1. \]  

(26)

where $F$ is the cumulative distribution function (cdf) of $X$.

- For a continuous random variable, $\xi_\alpha$ is the value such that

\[ F(\xi_\alpha) = \alpha. \]  

(27)
Sample quantiles

- For a sample $X_1, X_2, \ldots, X_T$, the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(T)}$ are the sample values ordered from smallest to largest. i.e.,

$$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(T)}. \quad (28)$$

- Then the $\alpha$–sample quantile is

$$\hat{\xi}_\alpha = X_{(m)}, \quad (29)$$

where

$$m = \lceil \alpha T \rceil, \quad (30)$$

and $\lceil x \rceil$ denotes the ceiling function, i.e., smallest integer not less than $x$.\(^7\)

\(^7\)We may also round to the nearest integer or interpolate between two order statistics.
Normal probability plot

- The $\alpha$–quantile of a normal distribution with mean $\mu$ and variance $\sigma^2$ is

$$\mu + \sigma \Phi^{-1}(\alpha), \quad (31)$$

where $\Phi^{-1}$ is the inverse of the standard normal cdf, i.e.,

$$\Phi^{-1}(\alpha) = z_\alpha \quad (32)$$

$$\Leftrightarrow \Phi(z_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_\alpha} e^{-x^2/2} dx = \alpha, \quad (33)$$

i.e., $z_\alpha$ is the $\alpha$-quantile of the standard normal distribution, e.g.,

$$\Phi^{-1}(0.025) \approx -1.96.$$

---

$^8$\(\Phi\) is the (almost universal) notation for the standard normal cdf (as is $\phi$ for its pdf).
Normal probability plot

• Thus, if our sample comes from a normal population, a plot of

\[ \Phi^{-1} \left( \frac{i}{T+1} \right) \]

against \( X(i), \ i = 1, \ldots, T. \) \hfill (34)

should approximately be linear.

• In (34), a divisor \( T + 1 \) instead of \( T \) is used to avoid \( \Phi^{-1}(1) = \infty \) for \( i = T. \)\(^9\)

• (To check against a distribution other than the normal, the normal inverse cdf in (34) has to be replaced accordingly by that of the hypothesized distribution.)

\(^{9}\)Sometimes the continuity correction \( \Phi^{-1} \left( \frac{i-0.5}{T} \right) \) is used instead.
**Shape of normal probability plots when the empirical distribution exhibits excess kurtosis**

- Excess kurtosis means that the probability of large negative or positive values is greater than under the corresponding normal density function.

- That is,
  - the lower quantiles are smaller than the normal quantiles, and
  - the upper quantiles are larger than the normal quantiles.

- Consequently, if the empirical quantiles are on the $x$–axis, and the theoretical ones are on the $y$–axis, then fat tails show up in an “$S$–shaped” QQ plot.

- That is, we observe deviations to the left of an ideal straight line in the left tail (lower quantiles), and to the right of an ideal straight line in the right tail (upper quantiles).
normal probability plot (qq plot) for the daily DAX 30 returns
normal probability plot (qq plot) for the daily S&P 500 returns
Moment–based measures of skewness and kurtosis

- Just as the first two moments (i.e., mean and variance) are often used to characterize location and spread (scale) of a distribution, (standardized) higher–order moments are frequently employed to characterize their shape.

- The standardized fourth moment is often used as a measure for excess kurtosis, i.e.,

\[ \kappa = \text{kurtosis}(r) = \frac{\mathbb{E}[(r - \mu)^4]}{\sigma^4} = \mathbb{E} \left[ \left( \frac{r - \mu}{\sigma} \right)^4 \right], \]  

(35)

where \( \mu \) and \( \sigma \) are the mean and standard deviation of \( r \), respectively.

- Note that, by standardization, this measure is independent of location (\( \mu \)) and scale (\( \sigma \)).
Moment–based kurtosis

- The moment–based kurtosis coefficient (standardized fourth moment)
  \[ \kappa = \frac{E[(r - \mu)^4]}{\sigma^4} = \mathbb{E} \left[ \left( \frac{r - \mu}{\sigma} \right)^4 \right] . \] (36)

- For normally distributed variables (irrespective of mean and variance)
  \[ \kappa_{\text{normal}} = 3, \] (37)
  and a leptokurtic distribution will have \( \kappa > 3 \).

- The intuition is that, due to the fourth power involved, the kurtosis measure is particularly sensitive to the weight of the tail regions far from the mean, so a higher \( \kappa \) indicates more probability mass in the tails.
Moment–based kurtosis

- For a sample \( r_1, r_2, \ldots, r_T \), we can estimate the kurtosis coefficient (35) via its sample analogue,

\[
\hat{\kappa} = \frac{T^{-1} \sum_{t=1}^{T} (r_t - \hat{\mu})^4}{\hat{\sigma}^4},
\]

where

\[
\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_t \quad \text{and} \quad \hat{\sigma} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_t - \bar{r})^2}
\]

are the sample mean and sample standard deviation, respectively.
Moment–based skewness

- Deviations from symmetry may also be present (and practically relevant), although these tend to be less pronounced and more difficult to detect.

- Roughly, left (negative) skewness means that the left tail of the distribution is longer than its right tail and vice versa for right (positive) skewness.

- The moment–based skewness measure of random variable $r$ is

$$s = \text{skewness}(r) = \frac{E[(r - \mu)^3]}{\sigma^3} = E \left[ \left( \frac{r - \mu}{\sigma} \right)^3 \right], \tag{40}$$

with sample counterpart

$$\hat{s} = \frac{T^{-1} \sum_{t=1}^{T} (r_t - \hat{\mu})^3}{\hat{\sigma}^3}. \tag{41}$$
Moment–based skewness

- In contrast to the kurtosis coefficient, the sign information is preserved in (40).

- For the normal (and any other symmetric density), positive and negative deviations from the mean cancel each other out, and so

\[ s = 0. \]

- On the other hand, if negative tail observations dominate, then \( s < 0 \), and the distribution is skewed to the left (*negative skewness*).

- If positive tail observations dominate, then \( s > 0 \), and the distribution is skewed to the right (*positive skewness*).
Moment–based measures of skewness and kurtosis: Not to be overinterpreted

- Moments are calculated by *averaging* over the entire support of a distribution.

- In general, although informative and useful (and routinely calculated), a single moment cannot always be taken as a trustworthy indicator of distributional shape, and may even be misleading at times.

- For example, a symmetric distribution always has $s = 0$, but $s = 0$ does not imply symmetry.
Moment–based measures of skewness and kurtosis: Not to be overinterpreted

• As a somewhat artificial example, consider the density (a mixture of normals)

\[ f(y) = \frac{4}{5}\phi(y; \sqrt{0.4}, 3) + \frac{1}{5}\phi(y; -\sqrt{6.4}, 1), \]  

(42)

where

\[ \phi(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\}. \]

• Density (42) has \( s = 0; \) the density is shown on the next slide.
An asymmetric density with moment-based skewness equal to zero
Moment–based measures of skewness and kurtosis

- Similar caveats apply to the kurtosis measure.

- Due to the higher–order moments involved, the moment–based measures of skewness and kurtosis are also rather sensitive to a few (even a single) large observations.\(^\text{10}\)


- In any case, a nonparametric density estimate will often be more informative.

\(^{10}\) I.e., have large sampling variance for fat–tailed distributions.
Jarque–Bera test for normality

- Measures $\hat{\kappa}$ and $\hat{s}$ can be used to construct a simple test for normality.

- Under normality,

$$\sqrt{T}\hat{s}/\sqrt{6} \xrightarrow{d} \text{Normal}(0, 1), \quad \text{and} \quad \sqrt{T}(\hat{\kappa} - 3)/\sqrt{24} \xrightarrow{d} \text{Normal}(0, 1),$$

and hence

$$T\hat{s}^2/6 \xrightarrow{d} \chi^2(1), \quad T(\hat{\kappa} - 3)^2/24 \xrightarrow{d} \chi^2(1). \quad (44)$$

- Moreover, both quantities in (44) are asymptotically independent, so the Jarque–Bera (JB) test statistic

$$JB = T\hat{s}^2/6 + T(\hat{\kappa} - 3)^2/24 \xrightarrow{d} \chi^2(2), \quad (45)$$

a $\chi^2$ distribution with two degrees of freedom.
Jarque–Bera test for normality

• Note that a separate test of

\[ H_0 : s = 0 \]

based on the asymptotic standard error \( \sqrt{\frac{6}{T}} \) of \( \hat{s} \) is still a test of normality (although not a recommendable one) and not a test of symmetry in general.

• The reason is that, although symmetric distributions always have \( s = 0 \), the asymptotic standard error \( \sqrt{\frac{6}{T}} \) is valid only under normality, and it can be much larger for fat–tailed symmetric distributions.\(^{11}\)

\(^{11}\) Or it may not even be well–defined in case the relevant moments are not finite.
Alternative Distributions for Returns

- Mandelbrot (1963), in a classic paper, was one of the first to provide a detailed study of the fat-tailedness of various speculative returns.\footnote{B. Mandelbrot (1963). The Variation of Certain Speculative Prices. \textit{Journal of Business} 36, 394-419. See, e.g., Mills and Markellos (2008) for a textbook discussion.}

- As an alternative distribution for returns, he suggested the \textit{stable Paretian}, or $\alpha$–stable, or Lévy stable, or just \textit{stable} distribution.

- We shall briefly consider \textit{symmetric} stable Paretian random variables.

- These can be extended to allow for asymmetries, but this would add little to the present discussion with its focus on fat-tailedness.
Alternative Distributions for Returns: Stable distributions

- A symmetric stable random variable $X$ is characterized by its characteristic exponent or index of stability, $\alpha$, with $0 < \alpha \leq 2$, which determines the degree of tail–fatness.

- Except for a few special cases, its density (pdf) is unavailable in closed–form.$^{13}$

- However, it can be defined via its characteristic function (cf), $\varphi_X(t; \alpha)$, given (for the location zero and unit scale version) by

$$\varphi_X(t; \alpha) = E(\exp\{itX\}) = \exp\{-|t|^\alpha\}, \quad 0 < \alpha \leq 2, \quad i = \sqrt{-1}. \quad (46)$$

---

$^{13}$But efficient methods exist for numerical computation.
Alternative Distributions for Returns: Stable distributions

- Note that, when $\alpha = 2$, (46) becomes

$$\phi_X(t; \alpha = 2) = \exp\{-t^2\},$$

which is equal to $\exp\{-t^2\sigma^2/2\}$ with $\sigma^2 = 2$, i.e., the cf of the normal distribution with variance 2.

- That is, the class of symmetric stable distributions nests the normal distribution for $\alpha = 2$.

- For $\alpha < 2$, however, the tail behavior is radically different from that of the Gaussian.

- Namely, explaining part of the name, its tail behavior for $0 < \alpha < 2$ is the same as that of the Pareto distribution, i.e., it has power tails with tail index $\alpha$. 
Power (or polynomial, or Pareto) tails

- One way to characterize a distribution is via its asymptotic tail behavior.

- E.g., the normal distribution has tails that decay to zero exponentially (i.e., rather fast).

- Thus it has finite moments of all orders, i.e., \( E(|X|^k) < \infty \) for all \( k > 0 \).

- Empirically, the tails of many financial variables appear to decay considerably slower.

- Namely, a polynomial, Pareto or power tail often provides a better description of the process generating large (positive and negative) returns.

- In this case, not all moments can exist finite.
Power (or polynomial, or Pareto) tails

- To illustrate, consider the (exact) *Pareto distribution* with cdf

\[
F(x; \alpha, x_0) = 1 - \left(\frac{x_0}{x}\right)^\alpha, \quad x > x_0 > 0, \quad \alpha > 0,
\]

(48)

and density

\[
f(x; \alpha, x_0) = \alpha x_0^\alpha x^{-(\alpha+1)}, \quad x > x_0.
\]

(49)

- The moments are given by (for \(m \neq \alpha\))

\[
E(X^m) = \alpha x_0^\alpha \int_{x_0}^{\infty} x^{-(\alpha+1-m)} \, dx
\]

(50)

\[
= \alpha x_0^\alpha \frac{x^{m-\alpha}}{m-\alpha} \bigg|_{x_0}^{\infty} = \begin{cases} \frac{\alpha x_0^m}{\alpha-m} & \text{for } m < \alpha \\ \infty & \text{for } m \geq \alpha \end{cases}.
\]

(51)
**Power (or polynomial, or Pareto) tails**

- The tails of the distribution decay so slowly that the improper integral does not converge for \( m \geq \alpha \).

- In many applications (e.g., risk management), the survivor function (the complimentary distribution function) of a random variable \( X \), defined by

\[
\overline{F}(x) = 1 - F(x) = \Pr(X > x)
\]  

is of interest, particularly as \( x \) becomes large.\(^{14} \)

- For \( X \sim \text{Pareto}(x_0, \alpha) \),

\[
\overline{F}(x) = cx^{-\alpha},
\]  

where \( c = x_0^\alpha \).

\(^{14}\)The same arguments can be made for the left tail, which is actually often of greater interest for risk management purposes.
Power (or polynomial, or Pareto) tails

- For a number of important distributions and financial models, the survivor function is asymptotically (i.e., as $x$ becomes large) of the form (53), i.e., somewhat informally,

$$
\bar{F}(x) \approx cx^{-\alpha} \quad \text{for large } x, \quad (54)
$$

for some constant $c > 0$.

- In this case, we say that the distribution has an asymptotic Pareto or power tail.\textsuperscript{15}

\textsuperscript{15}Sometimes the term fat tails is also reserved exclusively for power tails rather than for any leptokurtic distribution.
Power (or polynomial, or Pareto) tails

- The density is then asymptotically proportional to

\[ \frac{d(1 - cx^{-\alpha})}{dx} \propto x^{-(\alpha+1)}, \tag{55} \]

and it follows from the calculation above that moments exist only for orders below \( \alpha \).

- Quantity \( \alpha \) in (54) is referred to as the \textit{tail index}.

- The smaller \( \alpha \), the slower does the tail decay to zero (i.e., the fatter is the tail), and the more likely is the realization of an extreme event (as compared to the bulk of the data).

- E.g., for \( \alpha \leq 2 \) (\( \alpha \leq 1 \)) not even the variance (the mean) exists finite.
Power (or polynomial, or Pareto) tails: Comparison with exponential tails

- An exponential tail can be defined as

\[ F(x) \sim \exp(-x^b) \quad \text{as} \quad x \to \infty, \quad b > 0. \] (56)

- In (56), \( b < 2 \) implies tails which are thicker than those of the normal distribution.\(^{16}\)

- A power tail decays slower than any exponential tail, since

\[ \frac{\exp(-x^b)}{x^{-\alpha}} \xrightarrow{x \to \infty} 0 \quad \text{for all} \quad b > 0 \quad \text{and} \quad \alpha > 0. \] (57)

\(^{16}\)The case \( b < 1 \) is sometimes referred to as “stretched exponential”.

Power (or polynomial, or Pareto) tails: Comparison with exponential tails

- If tails are asymptotically exponential of the form (56), then

\[ \mathbb{E}(|X|^k) < \infty \quad \text{for all } k > 0, \quad (58) \]

i.e., all moments are finite.

- On the other hand, with asymptotic power tails of the form (53), we have seen that the tails decay so slowly that moments are finite only for order \( k < \alpha \),

\[ \mathbb{E}(|X|^k) = \infty \quad \text{for } k \geq \alpha. \quad (59) \]
Stable distributions (continued)

- For $0 < \alpha < 2$, the stable distribution has asymptotic power tails with index $\alpha$, i.e.,

$$\Pr(X > x) = \overline{F}(x) \simeq c(\alpha)x^{\alpha}, \quad \text{as } x \to \infty,$$

(60)

where $c(\alpha) = \sin(\pi \alpha/2) \Gamma(\alpha)/\pi$, and $\Gamma$ is the gamma function,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x} \, dx.$$  

(61)
Stable distributions: Tail behavior

- To illustrate the power–law property of the stable class for $0 < \alpha < 2$, we may consider another special case where the pdf is available in closed form, namely $\alpha = 1$.

- In this case we obtain the Cauchy distribution, which happens to belong to another well–known class of distributions, namely it is (also) the Student’s $t$ distribution with one degree of freedom.\(^{17}\)

- The density of the (location–zero scale–one) Cauchy distribution is

$$f(x) = \frac{1}{\pi} \frac{1}{1 + x^2}.$$ 

\(^{17}\)See Appendix.
Stable distributions: Tail behavior

• For large $x$, 
\[ f(x) = \frac{1}{\pi} \frac{1}{1 + x^2} \approx \frac{1}{\pi x^2}. \]  
(62)

• Since we obtain the pdf via differentiation of the cdf, for large $x$, 
\[ \overline{F}(x) = 1 - F(x) \approx \frac{1}{\pi x}, \]  
(63)

a power law with index $\alpha = 1$ (as in (60)).

• The Cauchy and the N(0,2) densities are shown on the next slide.
Stable distributions: Stability

• The power law decay of the tails (with tail index equal to its characteristic exponent) makes the stable distribution a potential candidate for modeling fat tailed phenomena such as financial returns.

• A further attractive property is its stability (under summation), explaining the other part of its name:

• Namely, sums of independent stable variables, each with the same tail index $\alpha$, also follow a stable distribution with index $\alpha$.

• This is well–known for the Gaussian from basic statistics, and stable distributions (which include the Gaussian) are the only ones with this property.
Stable distributions: Stability

• Let $Z$ have a location–zero scale–one symmetric stable distribution with characteristic function (46).

• Then variable

$$Y = \mu + cZ, \quad c > 0,$$

has a symmetric stable distribution with location $\mu$ and scale $c$,\(^{18}\) denoted as

$$Y \sim S_{\alpha}(\mu, c).$$

• (Note that $Z \sim S_{\alpha}(0, 1)$.)

• The characteristic function (cf) of $Y$ is

$$\varphi_Y(t) = \exp\{it\mu - c^\alpha|t|^\alpha\}.$$  

\(^{18}\)If $\alpha > 1$, then $\mu$ is the mean of $X$.  


Stable distributions: Stability

• Now let
  \[ Y_j \overset{\text{ind}}{\sim} S_\alpha(\mu_j, c_j), \quad j = 1, \ldots, n, \]
  where “ind” means that the \( Y_j \) are independent.

• The cf of \( S = \sum_{j=1}^{n} Y_j \) is
  \[
  \varphi_S(t) = \prod_{j=1}^{n} \exp \{ it\mu_j - c_j^\alpha |t|^\alpha \} = \exp \left\{ \begin{array}{l} \sum_{j=1}^{n} \mu_j - \sum_{j=1}^{n} c_j^\alpha |t|^\alpha \end{array} \right\}. \]
  \( (68) \)

• That is
  \[
  S \sim S_\alpha(\mu, c), \quad \text{where} \quad \mu = \sum_{j=1}^{n} \mu_j, \quad \text{and} \quad c = \left( \sum_{j=1}^{n} c_j^\alpha \right)^{1/\alpha}. \]
  \( (69) \)
Stable distributions: Generalized CLT

• Related to the stability property is the generalized central limit theorem (GLCT).

• Recall that the classical CLT states that (suitably scaled) sums of random variables converge to the Gaussian distribution, provided the variance is finite.

• Roughly, the generalized CLT allows to drop the assumption of finite variances, and it turns out that $\alpha$–stable distributions are the only possible limit distributions of suitably scaled sums.
Stable distributions: Generalized CLT

- It is in this sense that Fama (1965)\textsuperscript{19} argued that

\begin{quote}
"Mandelbrot’s hypothesis can actually be viewed as a generalization of the central–limit theorem arguments of Bachelier and Osborne to the case where the underlying distributions of price changes from transaction to transaction … have infinite variances."
\end{quote}

Stable distributions: Application to stock returns

- Fit a stable distribution of the form (65) to the daily S&P 500 and DAX 30 returns from January 2000 to October 2011 ($T = 3081$ daily return observations).

- Maximum likelihood estimates of the characteristic exponents are
  - $\hat{\alpha}_{S&P500} = 1.51$
  - $\hat{\alpha}_{DAX30} = 1.62$,

  indicating rather heavy tails.
Log-Density of S&P 500 returns

- empirical (kernel)
- fitted normal
- fitted stable
Density of DAX 30 returns

- empirical (kernel)
- fitted normal
- fitted stable
Problems with the Paretian stable distribution

• The (nonnormal) stable distribution implies a power tail with tail index $\alpha < 2$.

• That is, the tails decay so slowly that even the variance (second moment) does not exist.

• It has been argued that such tails are actually a bit too fat when compared to empirical tails.

• That is, a power law may often be appropriate as a description of the tail behavior, but with a tail index larger than 2.

• Also the implication of nonexisting variance may have undesirable consequences.\(^{20}\)

\(^{20}\)Which is of course not a reason to reject the stable distribution without consulting the data.
Asymptotic power tails: A quick look at the data

• A first informal check of the appropriateness of a power law (and a reasonable guess for the tail index) can be obtained by looking at a log–log plot of the sample (or empirical) complementary cdf,

\[ \overline{F}_T = 1 - F_T(x), \quad (70) \]

against \( x \) (for large \( x \)).

• That is, if

\[ 1 - F(x) \approx cx^{-\alpha} \quad \text{for large } x, \quad (71) \]

then

\[ \log(1 - F(x)) \approx \log c - \alpha \log x \quad \text{for large } x. \quad (72) \]

• Hence, a plot of the log of the empirical complementary cdf, \( 1 - F_T(x) \), against \( \log x \) (i.e., the logs of the ordered observations) should display a linear relation in the tail region.
Asymptotic power tails: A quick look at the data

- E.g., for the upper (right) tail, consider the ordered sample values (order statistics)
  \[ x(1) \leq x(2) \leq \cdots \leq x(T). \] (73)

- If we believe that the largest \( k \) observations belong to the upper tail region such that (71) holds, then we plot the pairs

\[
\left( \log x(j), \log \left[ 1 - F_T(x(j)) \right] \right)_{j=T-k+1, \ldots, T} = \left( \log x(j), \log \left[ 1 - \frac{j}{T+1} \right] \right)_{j=T-k+1, \ldots, T}. \] (74)
Asymptotic power tails: A quick look at the data

- Equation (72), using the tail observations (74), can also be fitted by OLS to obtain a rough estimate of the tail index $\alpha$.\textsuperscript{21}

- In general, when it comes to estimating the tail index, the choice of $k$ in (74) is crucial.

- Ideally, to keep the bias small, only observations from the tail region should be used, i.e., such that (71) is valid.

- Choosing $k$ too small will reduce the precision of the estimator, however.

- This is yet another bias/variance trade–off.

\textsuperscript{21}Better estimators exist for the tail index, but the question of where the tail region begins is always difficult. A Hill plot may occasionally help. For a detailed treatment of these questions, see, e.g., Embrechts/Klüppelberg/Mikosch: *Modelling Extremal Events for Insurance and Finance*, Springer.
Asymptotic power tails: A quick look at the data

- Power tail:

\[ \log(1 - F(x)) \approx \log c - \alpha \log x \quad \text{for large } x. \quad (75) \]

- Assuming approximate symmetry, we can pool both tails and consider the distribution of the absolute values (after demeaning).\(^{22}\)

- This is done on the next slide for the S&P 500 returns, which shows such a plot for the 10% tail of the absolute (demeaned) return observations, i.e., roughly the 300 largest absolute return observations.

- Also shown are the linear fit based on estimating (75) via OLS as well as the Gaussian fit.

\(^{22}\)Pooling negative and positive observations rather than treating the left and right tail separately increases the number of observations. In practice, one would want to test whether the assumption of homogeneity across the tails is justified.
10% tail of the S&P 500 returns (estimated $\alpha$: 2.918)
The Hill estimator for the tail index

• Hill (1975)\textsuperscript{23} derived a (frequently used) maximum–likelihood based estimator for the index of an asymptotic power tail.\textsuperscript{24}

• The order statistics

\[ x(1) \leq x(2) \leq \cdots \leq x(T). \tag{76} \]

• Suppose that the \( k \) largest observations \( x(T-k+1) \leq x(T-k+2) \leq \cdots \leq x(T) \) belong to the Pareto–type tail region.

• Then the Hill estimator (based on the \( k \) largest order statistics) is given by

\[
\hat{\alpha}_k = \left( \frac{1}{k-1} \sum_{j=1}^{k-1} \log x(T-j+1) - \log x(T-k+1) \right)^{-1}. \tag{77}
\]

\textsuperscript{23}A Simple General Approach to Inference About the Tail of a Distribution. \textit{Annals of Statistics} 3, 1163-1174.

\textsuperscript{24}For a textbook discussion, see Chapter 7.3 in Mills and Markellos (2008).
The Hill estimator for the tail index

- Hill estimator

\[
\hat{\alpha}_k = \left( \frac{1}{k-1} \sum_{j=1}^{k-1} \log x(T-j+1) - \log x(T-k+1) \right)^{-1}.
\]

- The Hill estimator is consistent under quite general conditions (provided \(k \to \infty\) as \(T \to \infty\) in an appropriate way\(^{25}\)).

- If the data are iid, then

\[
\sqrt{k}(\hat{\alpha}_k - \alpha) \xrightarrow{d} \text{N}(0, \alpha^2).
\] (78)

\(^{25}\)Such that \(k/T \to 0\) as \(T \to \infty\).
The Hill estimator for the tail index

• However, if the sample is not iid (e.g., due to GARCH effects), then the variance can be considerably larger than given in (78).\textsuperscript{26}

• Clearly the Hill estimator estimates something well-defined only if the data generating process is indeed characterized by a Pareto tail.

The Hill estimator for the tail index

• A crucial choice to be made when using the Hill estimator is the threshold value $k$, i.e., the number of order statistics to be used in estimation.

• Ideally, only observations from the tail region should be used.

• But choosing $k$ too small will reduce the precision of the estimator.

• A Hill plot may sometimes help, which is obtained by plotting

\[ \hat{\alpha}_k \text{ against } k. \]  \hspace{1cm} (79)

• If we can find a range of $k$–values where the estimate is approximately constant, this can be taken as a hint for where the tail index may be located.
Hill plot for the absolute (demeaned) S&P 500 returns

\[ \hat{\alpha}_k \]

number of observations used, \( k \)
Rationale for the Hill estimator

• Heuristically, suppose we have an iid sample of size $T$ from an exact Pareto distribution, with density

$$f(x) = \alpha u^\alpha x^{-(\alpha+1)}, \quad x \geq u > 0,$$

(80)

with $u$ known.

• By independence, the joint density of the sample is

$$f(x_1, x_2, \ldots, x_T) = \prod_{t=1}^{T} f(x_t) = \alpha^T u^{\alpha T} \left( \prod_{t=1}^{T} x_t \right)^{-(\alpha+1)}.$$  (81)
Rationale for the Hill estimator

• The log–likelihood function, $\ell(\alpha)$, is

$$
\ell(\alpha) = \log \prod_{t=1}^{T} f(x_t) 
$$

(82)

$$
= T \log \alpha + \alpha T \log u - (\alpha + 1) \sum_{t=1}^{T} \log x_t
$$

(83)

$$
= T \log \alpha - \alpha \sum_{t=1}^{T} (\log x_t - \log u) - \sum_{t=1}^{T} \log x_t.
$$

(84)

• Then the maximum likelihood estimator (MLE) is the solution of

$$
\ell'(\alpha) = \frac{T}{\alpha} - \sum_{t=1}^{T} (\log x_t - \log u) = 0.
$$

(85)
Rationale for the Hill estimator

- That is,
  \[
  \hat{\alpha}_{MLE} = \left( \frac{1}{T} \sum_{t=1}^{T} \log x_t - \log u \right)^{-1}.
  \]  \(86\)

- We believe only the tail area to behave Pareto-like, and thus we only select the \(k\) largest order statistics for estimation,
  \[
  x(T-k+1) \leq x(T-k+2) \leq \cdots \leq x(T).
  \]  \(87\)

- If the threshold value \(k\) has been selected wisely, then the power tail should roughly become effective for values larger than \(x(T-k+1)\).\(^{27}\)

- Thus it appears plausible that a reasonable estimator for the tail index \(\alpha\) is obtained by replacing \(u\) in \(86\) with \(x(T-k+1)\),\(^{28}\) which gives \((77)\).

\(^{27}\)Since ideally we use all and only the observations from the tail region.

\(^{28}\)Of course, in addition to using only the \(k\) largest order statistics in \((86)\).
Power tails

- It is now a widely held view that the distribution of asset returns can often be described as fat-tailed in the power law sense but with finite variance, with tail indices (e.g., for stock returns) typically falling in the region between 2.5 and 5.

- In this case, the variance is finite, and over longer horizons a central limit theorem effect will kick in and (longer–horizon) returns will approach the normal distribution.
Power tails: Interpretation

• We might argue that asset returns actually have finite support, implying finiteness of all moments and hence inappropriateness of a Pareto–type tail (for whatever tail index).

• However,
  "saying that the support of an empirical distribution is bounded says very little about the nature of outlier activity that may occur in the data."\(^{29}\)

• “[...] the point is that the series is *behaving* as if the moments do not exist.”\(^{30}\)


Power tails

- We will never know the “true” data generating process. For most practical applications,

“the relevant question is not to determine what is the true asymptotic tail, but what is the best effective description of the tails in the domain of useful applications.”

---

Student’s $t$ distribution

• The Student’s $t$ distribution in its standard version (as it appears in normal random sampling theory) has density

$$f(x; \nu) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma(\nu/2) \sqrt{\pi \nu}} \left\{ 1 + \frac{x^2}{\nu} \right\}^{-(\nu+1)/2}, \quad \nu > 0, \quad (88)$$

where $\nu$ is a shape parameter known as the degrees of freedom.

• In applications to financial returns, parameter $\nu > 0$ is estimated from the data to fit the shape of the return distribution.

• Parameter $\nu$ regulates the thickness of the tails.

• As $\nu$ becomes smaller, the tails become thicker.
Student’s $t$ distribution: Power tails

- The Student’s $t$ distribution has power tails with tail index $\nu$.

- To see this, somewhat informally, note that, for large $|x|$, the density is approximately proportional to

$$
\left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2} \approx \left(\frac{x^2}{\nu}\right)^{-(\nu+1)/2} \propto |x|^{-(\nu+1)},
$$

and thus the survivor function, for large $x$,

$$
1 - F(x) \approx cx^{-\nu}.
$$

- A smaller $\nu$ thus implies fatter tails.

- In contrast to the Paretian stable distribution, the tail index is not restricted to the range $(0, 2)$.

- Moments of the Student’s $t$ exist only for orders smaller than $\nu$. 
Student’s $t$ distribution

- As $\nu \to \infty$, normality is approached,

$$\lim_{\nu \to \infty} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} = \lim_{\nu \to \infty} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu}{2}}$$

$$= \left[ \left(1 + \frac{x^2}{\nu}\right)^{-\nu} \right]^{1/2}$$

$$= \left[ e^{-x^2} \right]$$

$$= \exp\{-x^2/2\}.$$
Student’s $t$ distribution: Moments

- The Student’s $t$ density (89) is symmetric, so all (existing) odd moments are zero.

- For $\nu > 2$, the variance of (89) is

$$\text{Var}(X) = \frac{\nu}{\nu - 2}, \quad (91)$$

and for $\nu > 4$ the kurtosis exists and is given by

$$\text{kurtosis}(X) = 3\frac{\nu - 2}{\nu - 4} = 3 + \frac{6}{\nu - 4}, \quad (92)$$

which approaches 3 as $\nu \to \infty$ (as expected).
Standardized Student’s $t$

- The *standardized* (rather than standard) Student’s $t$ distribution with unit variance (assuming $\nu > 2$) is often used in financial models.

- In view of (91), if $X$ is standard Student’s $t$ with density (88) and $\nu > 2$, then

$$Y = \sqrt{\frac{\nu - 2}{\nu}} X$$

has unit variance and density

$$f(y; \nu) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma(\nu/2)\sqrt{\pi(\nu - 2)}} \left\{ 1 + \frac{y^2}{\nu - 2} \right\}^{-(\nu+1)/2}, \quad \nu > 2. \quad (94)$$

- Clearly this scale transformation doesn’t affect any essential features of the distribution.
In applications, we will typically introduce location and scale parameters, and a variable with density

\[ f(y; \nu) = \frac{\Gamma \left( \frac{\nu+1}{2} \right)}{\Gamma(\nu/2) \sqrt{\pi(\nu - 2)\sigma}} \left\{ 1 + \frac{(y - \mu)^2}{(\nu - 2)\sigma^2} \right\}^{-(\nu+1)/2}, \quad \nu > 2, \]

has mean \( \mu \) and variance \( \sigma^2 \).

The next slide shows the densities of a few standardized Student’s \( t \) densities (they have all the same (unit) variance).
unit-variance Student-t densities

ν = 2.5
ν = 5
ν = 10
ν = 15
ν = ∞
unit-variance Student-t densities, log-scale

- $\nu = 2.5$
- $\nu = 5$
- $\nu = 10$
- $\nu = 15$
- $\nu = \infty$
Alternative distributions for returns: Student’s $t$

• In applications, parameter $\nu$ is jointly estimated with the location and scale parameters $\mu$ and $\sigma$, respectively (e.g., via maximum likelihood).

• Fitting the Student’s $t$ to the S&P 500 and DAX 30 returns, maximum likelihood estimates of the degrees of freedom parameter $\nu$ are

$$\hat{\nu}_{S&P500} = 2.6, \quad \text{and} \quad \hat{\nu}_{DAX} = 3.1,$$

respectively.

• The Student’s $t$ distribution is symmetric, but generalizations that allow for skewness exist (cf. skewed $t$ distributions).
Density of DAX 30 returns

- **Empirical (kernel)**
- **Fitted normal**
Density of DAX 30 returns

- empirical (kernel)
- fitted Student’s t
Log–Density of DAX 30 returns

- empirical (kernel)
- fitted normal
Log-Density of DAX 30 returns

- **empirical (kernel)**
- **fitted Student’s t**
normal probability plot (qq plot) for the daily S&P 500 returns
QQ plot for the daily S&P 500 returns

Student's t quantiles, $\nu = 2.5527$
normal probability plot (qq plot) for the daily DAX 30 returns

sample quantiles

standard normal quantiles

124
QQ plot for the daily DAX 30 returns

Student's t quantiles, $\nu = 3.1185$
Generalized Exponential Distribution (GED)

- The generalized exponential distribution\(^{32}\) (GED) has density

\[
f(x; p) = \frac{\lambda p}{2^{1/p+1} \Gamma(1/p)\sigma} \exp \left\{ -\frac{\lambda |x - \mu|^p}{2\sigma^p} \right\}, \quad p > 0, \tag{96}
\]

where \(\lambda = 2^{1/p} \sqrt{\Gamma(3/p)/\Gamma(1/p)}\).

- The density has mean \(\mu\), and it has been standardized such that it has variance \(\sigma^2\).

- The density is symmetric, and parameter \(p\) determines the thickness of the tails.

- Due to exponential tails, all moments are finite.

\(^{32}\)The name indicates that the exponent 2 in the normal density is generalized, not to the exponential distribution. It is also referred to as generalized error distribution or generalized power distribution.
Generalized Exponential Distribution (GED)

- The GED density:

\[
f(x; p) = \frac{\lambda p}{2^{1/p+1} \Gamma(1/p) \sigma} \exp \left\{ -\frac{\lambda^p |x - \mu|^p}{2 \sigma^p} \right\}, \quad p > 0.
\]

- \( p < 2 \) corresponds to leptokurtic distributions, whereas \( p > 2 \) corresponds to platykurtic (negative excess kurtosis) distributions.

- For \( p = 2 \), we have \( \lambda = 1 \) and we get the normal, and for \( p = 1 \) we obtain the Laplace (double exponential) distribution.

- As \( p \to \infty \), the GED approaches the uniform distribution.
Examples of zero-mean and unit-variance GED densities

- Laplace ($p = 1$)
- $p = 1.5$
- normal ($p = 2$)
- uniform ($p = \infty$)
Finite Normal Mixture Distributions

- Suppose returns $r_t$ are normally distributed but the mean and the variance may depend on the current state or regime of the market at time $t$.\textsuperscript{33}

- E.g., bull and bear market regimes.

- With $k$ regimes ($k = 2$ in the above example), we can write

$$r_t \sim \begin{cases} 
N(\mu_1, \sigma_1^2) & \text{if market is in Regime 1} \\
N(\mu_2, \sigma_2^2) & \text{if market is in Regime 2} \\
& \vdots \\
N(\mu_k, \sigma_k^2) & \text{if market is in Regime } k
\end{cases} \quad (97)$$

- Consider regime indicator $s_t$ such that $s_t = j$ if the market is in regime $j$ at time $t$, $j = 1, \ldots, k$.

\textsuperscript{33}That is, returns are normal conditional on the prevailing market regime.
Finite Normal Mixture Distributions

- Regimes are not observable and are not known \textit{ex ante}.

- However, we may be able to \textit{(ex ante)} assign probabilities to the market being in regime \( j \) at time \( t \), \( j = 1, \ldots, k \), i.e.,

\[
\Pr(s_t = j) =: \lambda_j, \quad \lambda_j \in (0, 1), \quad j = 1, \ldots, k, \quad \sum_{j=1}^{k} \lambda_j = 1. \tag{98}
\]

- Then we can write (97) as

\[
r_t \sim \begin{cases} 
N(\mu_1, \sigma_1^2) & \text{with probability } \lambda_1 \\
N(\mu_2, \sigma_2^2) & \text{with probability } \lambda_2 \\
\vdots & \\
N(\mu_k, \sigma_k^2) & \text{with probability } \lambda_k
\end{cases} \tag{99}
\]
Finite Normal Mixture Distributions

• With (99), and from the law of total probability, i.e., with exclusive and exhaustive events $B_i$, $i = 1, \ldots, n$,

$$
\Pr(A) = \sum_{i=1}^{n} \Pr(B_i)P(A|B_i), \quad (100)
$$

the distribution function (cdf) of $r_t$ is

$$
\Pr(r_t \leq r) = \sum_{j=1}^{k} \Pr(s_t = j) \Pr(r_t \leq r|s_t = j)
= \sum_{j=1}^{k} \lambda_j \Phi \left( \frac{r - \mu_j}{\sigma_j} \right), \quad (101)
$$

where the standard normal cdf

$$
\Phi(z) = (2\pi)^{-1/2} \int_{-\infty}^{z} e^{-\xi^2/2} d\xi. \quad (102)
$$
The density of $r_t$ is then

$$f(r; \theta) = \sum_{j=1}^{k} \Pr(s_t = j) f(r | s_t = j)$$

$$= \sum_{j=1}^{k} \lambda_j \frac{1}{\sigma_j} \phi \left( \frac{r - \mu_j}{\sigma_j} \right)$$

$$= \sum_{j=1}^{k} \lambda_j \phi(r; \mu_j, \sigma_j^2)$$

$$= \sum_{j=1}^{k} \lambda_j \phi(r; \mu_j, \sigma_j^2)$$

$$(\mu_i, \sigma_i^2) \neq (\mu_j, \sigma_j^2), \quad i \neq j, \quad (103)$$

where

$$\phi(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\},$$

and $\phi(x) = \phi(x; 0, 1)$ the standard normal pdf.
Finite Normal Mixture Distributions

- The \( k \)-component normal mixture density

\[
    f(r; \theta) = \sum_{j=1}^{k} \lambda_j f(r | s_t = j) = \sum_{j=1}^{k} \lambda_j \phi(r; \mu_j, \sigma^2_j), \tag{104}
\]

where the parameter vector (with \( 3k - 1 \) elements)

\[
    \theta = (\lambda_1, \ldots, \lambda_{k-1}, \mu_1, \ldots, \mu_k, \sigma^2_1, \ldots, \sigma^2_k)',
\]

with component densities

\[
    \phi(r; \mu_j, \sigma^2_j) = \frac{1}{\sqrt{2\pi\sigma_j}} \exp \left\{ -\frac{(r - \mu_j)^2}{2\sigma^2_j} \right\}, \quad j = 1, \ldots, k,
\]

component means \( \mu_j \), component variances \( \sigma^2_j \), and mixing weights \( \lambda_j \), \( j = 1, \ldots, k \).
Finite Normal Mixture Distributions

• Note that $r_t$ in this setting is not a linear combination of normal variables (which would itself be normal).

• Rather it is a random variable whose pdf is a linear combination of normal pdfs.

• The resulting distribution is not normal as long as $k > 1$.  

• In fact, normal mixture distributions are very flexible with respect to skewness and kurtosis (and further shape properties).  

---

34 And provided not all $(\mu_j, \sigma_j^2)$ are identical, which was excluded in (103).

35 Indeed, they are often and in many diverse areas used just as a flexible modeling tool without mapping any intuition to the model structure (such as the regime interpretation referred to above).
• To illustrate, consider the *scale mixture* case, where

\[ \mu_1 = \mu_2 = \cdots = \mu_k = \mu \quad (= \text{E}(r_t)) \quad (105) \]

(i.e., all means are identical, we only have different volatility regimes).

• Then the density is symmetric but it exhibits excess kurtosis relative to the normal.
Finite Normal Mixture Distributions: Scale Mixtures

• To see the intuition behind this fact, consider a two–component scale mixture with

$$\lambda_1 = 0.8, \quad \lambda_2 = 1 - \lambda_1 = 0.2, \quad \sigma_1^2 = 0.5, \quad \sigma_2^2 = 8,$$

which has variance\(^\text{36}\)

$$\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2 = 0.8 \times 0.5 + 0.2 \times 8 = 0.4 + 1.6 = 2.$$  \hspace{1cm} (107)

• In the long run, 80% of the returns are generated by Regime 1 with low variance, accounting for a higher peak around the mean as compared to a normal density with the same (overall) variance (107).

• 20% of the returns are generated by Regime 2 with very high variance (as compared to the average of 2), generating more pronounced tails as compared to a normal density with variance 2.

\(^{36}\text{Note: The variance formula for the case with different regime means } (\mu_1 \neq \mu_2) \text{ is more complicated than (107).}\)
\[ r_t \sim \mathcal{N}(0, 2) \]
$r_t \sim \text{Normal}(0, 2)$
$r_t \sim \text{Normal} \left( 0, \frac{1}{2} \right) \text{ with prob. 0.8}$
\( r_t \sim \begin{cases} \text{Normal}(0, \frac{1}{2}) & \text{with prob. 0.8} \\ \text{Normal}(0, 8) & \text{with prob. 0.2} \end{cases} \)
\[ r_t \sim \begin{cases} 
\text{Normal}(0, \frac{1}{2}) & \text{with prob. 0.8} \\
\text{Normal}(0, 8) & \text{with prob. 0.2}
\end{cases} \Rightarrow f(r_t) = 0.8 \cdot N\left(0, \frac{1}{2}\right) + 0.2 \cdot N(0, 8) \]
Finite Normal Mixture Distributions: Skewness

- A skewed (asymmetric) density can be generated by allowing the component means to differ.

- E.g., for stock index returns the typical situation where left-skewness emerges is where the component with the smaller probability (mixing weight) has the greater variance and the smaller mean (so the left tail is flattened out).
Alternative Distributions for Returns: Discrete Normal Mixtures

- For the S&P 500 and the DAX, (maximum likelihood) parameter estimates are

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\lambda}_1$</th>
<th>$\hat{\mu}_1$</th>
<th>$\hat{\sigma}_1^2$</th>
<th>$\hat{\lambda}_2$</th>
<th>$\hat{\mu}_2$</th>
<th>$\hat{\sigma}_2^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S&amp;P 500</td>
<td>0.718</td>
<td>0.062</td>
<td>0.530</td>
<td>0.282</td>
<td>-0.154</td>
<td>5.158</td>
</tr>
<tr>
<td>DAX 30</td>
<td>0.739</td>
<td>0.071</td>
<td>1.017</td>
<td>0.261</td>
<td>-0.215</td>
<td>7.156</td>
</tr>
</tbody>
</table>

Typically you would provide standard errors along with parameter estimates.
Density of S&P 500 returns

- Blue line: empirical (kernel)
- Red dashed line: fitted normal mixture (k = 2)
Log–Density of S&P 500 returns

- empirical (kernel)
- fitted normal mixture (k = 2)
Discrete Normal Mixtures: Extensions

- The basic mixture specification can be generalized in various directions to provide a more complete return model.

- For example, the mixing weights $\lambda_j$ can be made time–varying, so that the regimes are persistent or depend on a set of predetermined variables (and are predictable).
Discrete Normal Mixtures: Extensions

• E.g., consider persistent regimes:

• This means that the probability that a high–volatility regime is followed by a high–volatility regime is larger than the unconditional (long–run) probability of a high–volatility regime.  

• In terms of conditional probabilities, regime $j$ is persistent if

$$\Pr(s_t = j | s_{t-1} = j) > \Pr(s_t = j).$$

(108)

• If the conditional regime probabilities at time $t$ are allowed to depend on the prevailing regime at time $t - 1$ (as in (108)), then the basic mixture model becomes a Markov–switching model, which is a very popular framework for modeling persistent regimes.

---

38The same applies to low–volatility or any other type of regimes.
Discrete Normal Mixtures: Extensions

- If regimes are persistent, then high–volatility regimes (and low–volatility regimes as well) will be clustered in time, giving rise to volatility clustering, our next stylized fact.
Temporal Properties of Returns

S&P 500 index returns (daily), January 2000 to October 2011
Temporal Properties of Returns

• Observe time series of returns $r_1, r_2, \ldots, r_T$.

• Consider the sample autocorrelation function (SACF) at lag $\tau$,

$$
\hat{\rho}(\tau) = \frac{\sum_{t=1}^{T-\tau} (r_t - \bar{r})(r_{t+\tau} - \bar{r})}{\sum_{t=1}^{T} (r_t - \bar{r})^2}, \quad \tau > 0, \quad (109)
$$

where

$$
\bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t, \quad (110)
$$

and $T$ is the sample size.

• SACF (109) can also be computed for the squared or absolute returns.
Temporal Properties of Returns

- The SACF is often plotted along with (frequently 95%) confidence intervals around zero.

- These are based on the fact that, if the process under study is strict (independent) white noise, then (under some further moment condition)\(^{39}\)

\[ \sqrt{T} \hat{\rho}(\tau) \xrightarrow{d} N(0,1), \quad \tau > 0, \quad (111) \]

and the sample autocorrelations at different lags \(\tau\) are asymptotically independent.

Temporal Properties of Returns

• Thus we can treat \( \hat{\rho}(\tau) \) as asymptotically normal with variance \( 1/T \) under the hypothesis of a strict white noise process, i.e.,

\[
\hat{\rho}(\tau) \xrightarrow{\text{asym}} N\left(0, \frac{1}{T}\right),
\]

\((112)\)

and the 95% intervals are given by

\[
\pm \frac{1.96}{\sqrt{T}}, \quad (1.96 = 0.975–\text{quantile of the standard normal}). \quad (113)
\]
- Dashed lines represent 95% asymptotic confidence intervals associated with a strict white noise process:
autocorrelations of absolute DAX 30 returns

autocorrelations of squared DAX 30 returns
Temporal Properties of Returns

- Return series are characterized by *volatility clustering*, that is, “large [price] changes tend to be followed by large changes—of either sign—and small changes tend to be followed by small changes”.\(^{40}\)

- Thus variance (and thus risk) appears to be persistent and predictable.

- This is in contrast to the direction of price changes: Inspection of the SACF of the raw (untransformed) returns shows that these are basically unpredictable.

- One of the most popular approaches for capturing time-varying conditional volatility in financial returns is the GARCH model.

Properties of Financial Time Series

• The marginal distribution is not normal (in particular, fat tails, excess kurtosis).

• Volatility clustering.

• These properties relate to the univariate distribution of returns.
Dependence Structure of Returns

• In portfolio management, the dependence structure of returns is of crucial importance.

• E.g., in the classical Markowitz ($\mu - \sigma$) portfolio theory, we are interested in the first two moments of the portfolio return distribution, i.e., mean (i.e., expected return) and variance (risk):
  – For a given portfolio mean, investors prefer a lower variance (less risk).
  – For a given portfolio variance, investors prefer a higher expected return.

• In this framework, correlations between assets are of predominant interest, because the strength of the correlations determines the degree of risk (variance) reduction that can be achieved by portfolio diversification.
Dependence Structure of Returns

• To illustrate, suppose we just have two assets with (discrete) returns $R_1$ and $R_2$, respectively, with expected returns and variances

\[
\begin{align*}
E(R_1) &= \mu_1, \quad \text{Var}(R_1) = \sigma_1^2 \\
E(R_2) &= \mu_2, \quad \text{Var}(R_2) = \sigma_2^2.
\end{align*}
\]

• The covariance between $R_1$ and $R_2$ is

\[
\sigma_{12} = \text{Cov}(R_1, R_2) = E\{(R_1 - \mu_1)(R_2 - \mu_2)\} = \rho_{12}\sigma_1\sigma_2, \quad \text{(114)}
\]

where $\rho_{12}$ is the coefficient of correlation,

\[
\rho_{12} = \frac{\text{Cov}(R_1, R_2)}{\sqrt{\text{Var}(R_1)}\sqrt{\text{Var}(R_2)}} = \frac{\sigma_{12}}{\sigma_1\sigma_2}. \quad \text{(115)}
\]
Dependence Structure of Returns

- Now suppose the fraction of wealth invested in the first asset (i.e., the portfolio weight of Asset 1) is $x$.

- Then the portfolio return $R_p$ is

$$R_p = xR_1 + (1 - x)R_2. \quad (116)$$

- The expected portfolio return is

$$\mu_p = \mathbb{E}(R_p) = x \mathbb{E}(R_1) + (1 - x) \mathbb{E}(R_2) = x \mu_1 + (1 - x) \mu_2. \quad (117)$$

- The portfolio variance $\sigma_p^2$ (variance of the portfolio return) is

$$\sigma_p^2 = \text{Var}(xR_1 + (1 - x)R_2)$$
$$= x^2 \text{Var}(R_1) + (1 - x)^2 \text{Var}(R_2) + 2x(1 - x) \text{Cov}(R_1, R_2)$$
$$= x^2 \sigma_1^2 + (1 - x)^2 \sigma_2^2 + 2x(1 - x) \sigma_{12}$$
$$= x^2 \sigma_1^2 + (1 - x)^2 \sigma_2^2 + 2x(1 - x) \rho_{12} \sigma_1 \sigma_2.$$
Dependence Structure of Returns

- For each $x$, we obtain a specific $(\sigma_p, \mu_p)$-combination, and investors will chose the portfolio (i.e., $x$) which best suits their risk/expected return–preferences.

- In the two–asset example, the set of feasible $(\sigma_p, \mu_p)$–combinations can be worked out just by varying $x$ over the range of admissible values.

- (E.g., if short sales are prohibited, then $0 \leq x \leq 1$.)
Dependence Structure of Returns

• The set of possible \((\sigma_p, \mu_p)\)–combinations is typically represented in a figure with the portfolio standard deviation \((\sigma_p)\) of the \(x\)–axis and the portfolio mean \((\mu_p)\) on the \(y\)–axis.

• Consider an example with

\[
\mu_1 = 0.75, \quad \sigma_1^2 = 33 \ (\sigma_1 \approx 5.74), \quad \mu_2 = 1, \quad \sigma_2^2 = 35 \ (\sigma_2 \approx 5.92),
\]

and a couple of values for \(\rho_{12}\).
portfolio standard deviation, $\sigma_p$

portfolio mean, $\mu_p$

$\rho = -1$
$\rho = -0.5$
$\rho = 0$
$\rho = 0.25$
$\rho = 0.5$
$\rho = 0.75$
$\rho = 1$
Dependence Structure of Returns

- The degree of feasible diversification (risk reduction) strongly depends on $\rho_{12}$ as a measure for the strength of dependence between $R_1$ and $R_2$.

- This is intuitively clear, since correlations measure the tendency of assets to rise and fall together.

- Thus high correlations indicate that there is little potential for portfolio diversification and thus, e.g., protection against large losses in bearish markets.
Dependence Structure of Returns

- In practice, correlations are unknown and have to be estimated.

- With a sample of historical data of size $T$,

  \[(R_{1t}, R_{2t}), \quad t = 1, \ldots, T,\]  

  an obvious estimator is the sample correlation, i.e.,

  \[
  \hat{\rho}_{12} = \frac{\sum_{t=1}^{T} (R_{1t} - \overline{R}_1)(R_{2t} - \overline{R}_2)}{\sqrt{\sum_{t=1}^{T} (R_{1t} - \overline{R}_1)^2} \sqrt{\sum_{t=1}^{T} (R_{2t} - \overline{R}_2)^2}},
  \]

  where

  \[
  \overline{R}_1 = \frac{1}{T} \sum_{t=1}^{T} R_{1t}, \quad \overline{R}_2 = \frac{1}{T} \sum_{t=1}^{T} R_{2t}.
  \]
Dependence Structure of Returns

- Simple correlation estimates (averaging over different market periods) may be misleading, however, due to \textit{asymmetric dependence structures}.

- This refers to the observation that, for example, stock returns are more dependent in bear markets (market downturns) than in bull markets.

- Therefore, diversification might fail exactly when the benefits from diversification are most urgently needed.
Dependence Structure of Returns

- A popular tool to detect this asymmetric dependence structure are the **exceedance correlations** introduced by Longin and Solnik (2001).\(^{41}\)

- For a given threshold \(\theta\), the exceedance correlation between (demeaned) returns \(R_1\) and \(R_2\) is given by

\[
\rho(\theta) = \begin{cases} 
\text{Corr}(R_1, R_2|R_1 > \theta, R_2 > \theta) & \text{for } \theta \geq 0 \\
\text{Corr}(R_1, R_2|R_1 < \theta, R_2 < \theta) & \text{for } \theta \leq 0 
\end{cases}
\]  

(121)

- That is, for a given \(\theta > 0\) (\(\theta < 0\)), \(\rho(\theta)\) is the correlation between \(R_1\) and \(R_2\) conditional on the event that both returns are larger (smaller) than \(\theta\).

- Note that there are two exceedance correlations for \(\theta = 0\): One for \(\{R_1 < 0, R_2 < 0\}\) and one for \(\{R_1 > 0, R_2 > 0\}\).

Let us consider monthly returns of MSCI stock market indices for the US and Germany from January 1970 to October 2011 ($T = 501$ monthly observations).
• Sample exceedance correlations for a threshold of $\theta = -5\%$ can be calculated by picking the (bivariate) observations colored red.
Dependence Structure of Returns

- Theoretical exceedance correlations as implied by a time-invariant bivariate normal distribution can be calculated analytically or obtained by simulation.

- What would we expect them to look like for a time-invariant normal distribution?

- Of course they must be symmetric.
Exceedance correlations for the pair Germany/USA (MSCI stock market returns)

\[ \text{exceedance correlation } \rho(\theta) \]

- Data
- Normal distribution

Exceedance correlations for the pair Germany/USA (MSCI stock market returns)
Exceedance correlations for the pair Germany/UK (MSCI stock market returns)

- Exceedance correlation $\rho(\theta)$
- Data
- Normal distribution

The graph shows the exceedance correlation $\rho(\theta)$ for the pair Germany/UK (MSCI stock market returns). The data is represented by a black asterisk (*) line, while the normal distribution is shown by an orange circle (o) line.
Appendix: Symmetric stable distribution for $\alpha = 1$

- With $\alpha = 1$, the characteristic function (46) becomes $e^{-|t|}$, and the density can be calculated via the inversion formula,

\[
2\pi f(x) = \int_{-\infty}^{\infty} e^{-itx} e^{-|t|} dt
\]

\[
= \int_{-\infty}^{0} \cos(tx)e^t dt + \int_{0}^{\infty} \cos(tx)e^{-t} dt
\]

\[
- i \int_{-\infty}^{0} \sin(tx)e^t dt - i \int_{0}^{\infty} \sin(tx)e^{-t} dt
\]

\[
= 2 \int_{0}^{\infty} \cos(tx)e^{-t} dt,
\]

where the last line follows from $\cos(x) = \cos(-x)$ and $\sin(x) = -\sin(-x)$.
Appendix: Symmetric stable distribution for $\alpha = 1$

• Integration by parts twice then shows

$$f(x) = \frac{1}{\pi} \int_0^\infty \cos(tx)e^{-t}dt$$

$$= \frac{1}{\pi} \frac{1}{1 + x^2},$$

which is the location–zero scale–one Cauchy density.