

Statistical Tools in Finance and Insurance

Pavel Čížek, Wolfgang Härdle, Rafal Weron

November 25, 2003

Contents

I	Finance	9
1	Stable distributions in finance	11
	<i>Szymon Borak, Wolfgang Härdle, Rafal Weron</i>	
1.1	Introduction	11
1.2	α -stable distributions	12
1.2.1	Characteristic function representation	14
1.2.2	Simulation of α -stable variables	16
1.2.3	Tail behavior	18
1.3	Estimation of parameters	18
1.3.1	Tail exponent estimation	19
1.3.2	Sample Quantiles Methods	22
1.3.3	Sample Characteristic Function Methods	23
1.4	Financial applications of α -stable laws	26
2	Tail dependence	33
	<i>Rafael Schmidt</i>	
2.1	Tail dependence and copulae	33
2.2	Calculating the tail-dependence coefficient	36

2.2.1	Archimedean copulae	36
2.2.2	Elliptically contoured distributions	37
2.2.3	Other copulae	40
2.3	Estimating the tail-dependence coefficient	43
2.4	Estimation and empirical results	45
	Bibliography	53
3	Implied Trinomial Trees	55
	<i>Karel Komorád</i>	
3.1	Introduction	55
3.2	Basic Option Pricing Overview	57
3.3	Trees and Implied Models	59
3.4	ITT's and Their Construction	62
3.4.1	Basic insight	62
3.4.2	State space	64
3.4.3	Transition probabilities	66
3.4.4	Possible pitfalls	67
3.4.5	Illustrative examples	68
3.5	Computing Implied Trinomial Trees	74
3.5.1	Basic skills	74
3.5.2	Advanced features	81
3.5.3	What is hidden	84
	Bibliography	87
4	Functional data analysis	89
	<i>Michal Benko, Wolfgang Härdle</i>	
4.1	Introduction	89

5 Nonparametric Productivity Analysis 91*Wolfgang Härdle, Seok-Oh Jeong*

5.1	Introduction	91
5.2	Nonparametric Hull Methods	93
5.2.1	An Overview	93
5.2.2	Data Envelopment Analysis	94
5.2.3	Free Disposal Hull	94
5.3	DEA in Practice : Insurance Agencies	95
5.4	FDH in Practice : Manufacturing Industry	96

6 Money Demand Modelling 103*Noer Azam Achsani, Oliver Holtemöller and Hizir Sofyan*

6.1	Introduction	103
6.2	Money Demand	104
6.2.1	General Remarks and Literature	104
6.2.2	Econometric Specification of Money Demand Functions	105
6.2.3	Estimation of Indonesian Money Demand	108
6.3	Fuzzy Model Identification	113
6.3.1	Fuzzy Clustering	113
6.3.2	Takagi-Sugeno Approach	114
6.3.3	Model Identification	115
6.3.4	Modelling Indonesian Money Demand	117
6.4	Conclusions	118
	Bibliography	121

7 The exact LR test of the scale in the gamma family 125*Milan Stehlík*

7.1	Introduction	125
7.2	Computation the exact tests in the XploRe	127
7.3	Illustrative examples	128
7.3.1	Time processing estimation	128
7.3.2	Estimation with missing time-to-failure information	132
7.4	Implementation to the XploRe	137
7.5	Asymptotical optimality	138
7.6	Information and exact testing in the gamma family	139
7.7	The Lambert W function	140
7.8	Oversizing of the asymptotics	141
	Bibliography	143
8	Pricing of catastrophe (CAT) bonds	147
	<i>Krzysztof Burnecki, Grzegorz Kukla, David Taylor</i>	
8.1	Introduction	147
9	Extreme value theory	149
	<i>Krzysztof Jajuga, Daniel Papla</i>	
9.1	Introduction	149
10	Applying Heston's stochastic volatility model to FX options markets	151
	<i>Uwe Wystup, Rafal Weron</i>	
10.1	Introduction	151
11	Mortgage backed securities: how far from optimality	153
	<i>Nicolas Gaussel, Julien Tamine</i>	
11.1	Introduction	153
12	Correlated asset risk and option pricing	155

Wolfgang Härdle, Matthias Fengler, Marc Tisserand

12.1 Introduction	155
-----------------------------	-----

II Insurance 157

13 Loss distributions 159

Krzysztof Burnecki, Grzegorz Kukła, Rafal Weron

13.1 Introduction	159
-----------------------------	-----

14 Visualization of the risk process 161

Pawel Mista, Rafal Weron

14.1 Introduction	161
-----------------------------	-----

15 Approximation of ruin probability 163

Krzysztof Burnecki, Pawel Mista, Aleksander Weron

15.1 Introduction	163
-----------------------------	-----

16 Deductibles 165

Krzysztof Burnecki, Joanna Nowicka-Zagrajek, Aleksander Weron, A. Wyłomańska

16.1 Introduction	165
-----------------------------	-----

17 Premium calculation 167

Krzysztof Burnecki, Joanna Nowicka-Zagrajek, W. Otto

17.1 Introduction	167
-----------------------------	-----

18 Premium calculation when independency and normality assumptions are relaxed 169

W. Otto

18.1 Introduction	169
-----------------------------	-----

19 Joint decisions on premiums, capital invested in insurance company, rate of return on that capital and reinsurance	171
<i>W. Otto</i>	
19.1 Introduction	171
20 Stable Levy motion approximation in collective risk theory	173
<i>Hansjoerg Furrer, Zbigniew Michna, Aleksander Weron</i>	
20.1 Introduction	173
21 Diffusion approximations in risk theory	175
<i>Zbigniew Michna</i>	
21.1 Introduction	175

Part I

Finance

1 Stable distributions in finance

Szymon Borak, Wolfgang Härdle, Rafał Weron

1.1 Introduction

Stable laws – also called α -stable or Levy-stable – are a rich family of probability distributions that allow skewness and heavy tails and have many interesting mathematical properties. They appear in the context of the Generalized Central Limit Theorem which states that the only possible non-trivial limit of normalized sums of independent identically distributed variables is α -stable. The Standard Central Limit Theorem states that the limit of normalized sums of independent identically distributed terms with finite variance is Gaussian (α -stable with $\alpha = 2$).

It is often argued that financial asset returns are the cumulative outcome of a vast number of pieces of information and individual decisions arriving almost continuously in time (McCulloch, 1996; Rachev and Mitnik, 2000). Hence, it is natural to consider stable distributions as approximations. The Gaussian law is by far the most well known and analytically tractable stable distribution and for these and practical reasons it has been routinely postulated to govern asset returns. However, financial asset returns are usually much more leptokurtic, i.e. have much heavier tails. This leads to considering the non-Gaussian ($\alpha < 2$) stable laws, as first postulated by Benoit Mandelbrot in the early 1960s (Mandelbrot, 1997).

Apart from empirical findings, in some cases there are solid theoretical reasons for expecting a non-Gaussian α -stable model. For example, emission of particles from a point source of radiation yields the Cauchy distribution ($\alpha = 1$), hitting times for Brownian motion yield the Levy distribution ($\alpha = 0.5, \beta = 1$), the gravitational field of stars yields the Holtsmark distribution ($\alpha = 1.5$), for a review see Janicki and Weron (1994) or Uchaikin and Zolotarev (1999).

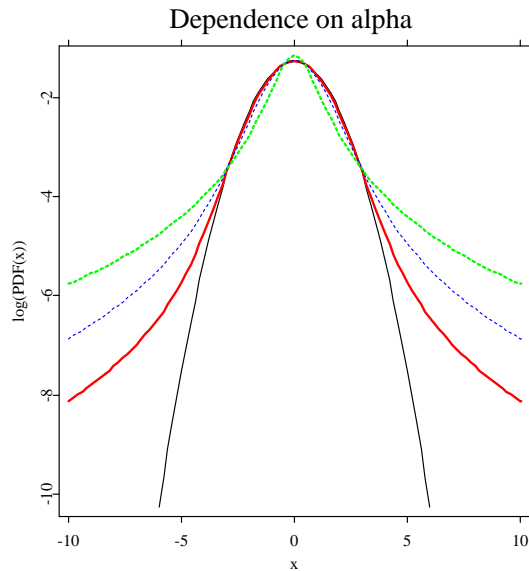


Figure 1.1: A semilog plot of symmetric ($\beta = \mu = 0$) α -stable probability density functions for $\alpha = 2$ (thin black), 1.8 (red), 1.5 (thin, dashed blue) and 1 (dashed green). The Gaussian ($\alpha = 2$) density forms a parabola and is the only α -stable density with exponential tails.

STFstab01.xpl

1.2 α -stable distributions

Stable laws were introduced by Paul Levy during his investigations of the behavior of sums of independent random variables in the early 1920s (Levy, 1925). A sum of two independent random variables having an α -stable distribution with index α is again α -stable with the same index α . This invariance property does not hold for different α 's, i.e. a sum of two independent stable random variables with different α 's is not α -stable. However, it is fulfilled for a more general class of infinitely divisible distributions, which are the limiting laws for sums of independent (but not identically distributed) variables.

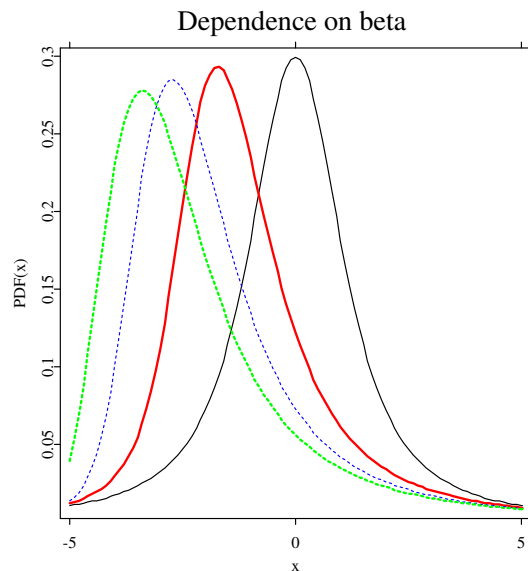


Figure 1.2: α -stable probability density functions for $\alpha = 1.2$ and $\beta = 0$ (thin black), 0.5 (red), 0.8 (thin, dashed blue) and 1 (dashed green).

STFstab02.xpl

The α -stable distribution requires four parameters for complete description: an index of stability $\alpha \in (0, 2]$ also called the tail index, tail exponent or characteristic exponent, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\sigma > 0$ and a location parameter $\mu \in \mathbb{R}$. The tail exponent α determines the rate at which the tails of the distribution taper off, see Figure 1.1. When $\alpha = 2$, a Gaussian distribution results. When $\alpha < 2$, the variance is infinite. When $\alpha > 1$, the mean of the distribution exists and is equal to μ . In general, the p th moment of a stable random variable is finite if and only if $p < \alpha$. When the skewness parameter β is positive, the distribution is skewed to the right, i.e. the right tail is thicker, see Figure 1.2. When it is negative, it is skewed to the left. When $\beta = 0$, the distribution is symmetric about μ . As α approaches 2, β loses its effect and the distribution approaches the Gaussian distribution

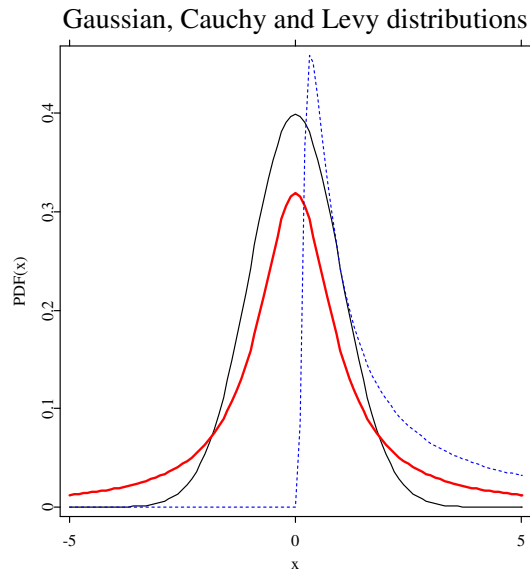


Figure 1.3: Closed form formulas for densities are known only for three distributions: Gaussian ($\alpha = 2$; thin black), Cauchy ($\alpha = 1$; red) and Levy ($\alpha = 0.5, \beta = 1$; thin, dashed blue). The latter is a totally skewed distribution, i.e. its support is \mathbb{R}_+ . In general, for $\alpha < 1$ and $\beta = 1$ (-1) the distribution is totally skewed to the right (left).

STFstab03.xpl

regardless of β . The last two parameters, σ and μ , are the usual scale and location parameters, i.e. σ determines the width and μ the shift of the mode (the peak) of the distribution.

1.2.1 Characteristic function representation

Due to the lack of closed form formulas for densities for all but three distributions (see Figure 1.3), the α -stable law can be most conveniently described by its characteristic function $\phi(t)$ – the inverse Fourier transform of the prob-

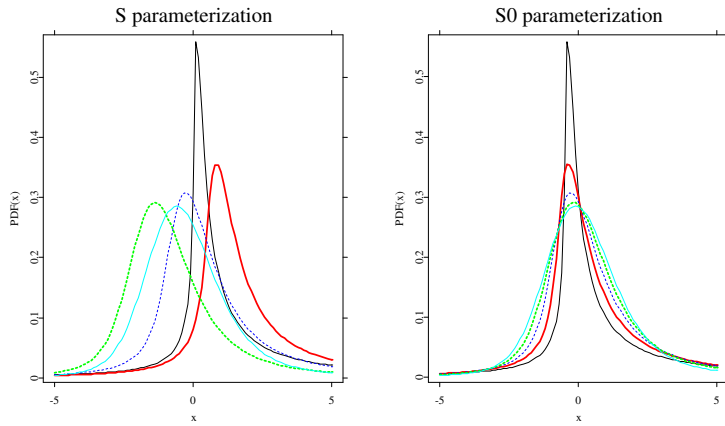


Figure 1.4: Comparison of S and S^0 parameterizations: α -stable probability density functions for $\beta = 0.5$ and $\alpha = 0.5$ (thin black), 0.75 (red), 1 (thin, dashed blue), 1.25 (dashed green) and 1.5 (thin cyan).

■ STFstab04.xpl

ability density function. However, there are multiple parameterizations for α -stable laws and much confusion has been caused by these different representations, see Figure 1.4. The variety of formulas is caused by a combination of historical evolution and the numerous problems that have been analyzed using specialized forms of the stable distributions. The most popular parameterization of the characteristic function of $X \sim S_\alpha(\sigma, \beta, \mu)$, i.e. an α -stable random variable with parameters α , σ , β and μ , is given by (Samorodnitsky and Taqqu, 1994; Weron, 1996):

$$\log \phi(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \{1 - i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2}\} + i\mu t, & \alpha \neq 1, \\ -\sigma |t| \{1 + i\beta \operatorname{sign}(t) \frac{2}{\pi} \log |t|\} + i\mu t, & \alpha = 1. \end{cases} \quad (1.1)$$

For numerical purposes, it is often useful (Fofack and Nolan, 1999) to use a different parameterization:

$$\log \phi_0(t) = \begin{cases} -\sigma^\alpha |t|^\alpha \{1 + i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} [(\sigma|t|)^{1-\alpha} - 1]\} + i\mu_0 t, & \alpha \neq 1, \\ -\sigma |t| \{1 + i\beta \operatorname{sign}(t) \frac{2}{\pi} \log(\sigma|t|)\} + i\mu_0 t, & \alpha = 1. \end{cases} \quad (1.2)$$

The $S_\alpha^0(\sigma, \beta, \mu_0)$ parameterization is a variant of Zolotariev's (M)-parameterization (Zolotarev, 1986), with the characteristic function and hence the density and the distribution function jointly continuous in all four parameters, see Figure 1.4. In particular, percentiles and convergence to the power-law tail vary in a continuous way as α and β vary. The location parameters of the two representations are related by $\mu = \mu_0 - \beta\sigma \tan \frac{\pi\alpha}{2}$ for $\alpha \neq 1$ and $\mu = \mu_0 - \beta\sigma \frac{2}{\pi} \log \sigma$ for $\alpha = 1$.

The probability density function and the cumulative distribution function of α -stable random variables can be easily calculated in XploRe. Quantlets `pdfstab` and `cdfstab` compute the pdf and the cdf, respectively, for a vector of values \mathbf{x} with given parameters `alpha`, `sigma`, `beta`, and `mu`, and an accuracy parameter `n`. Both quantlets utilize Nolan's (1997) integral formulas for the density and the cumulative distribution function. The larger the value of `n` (default `n=2000`) the more accurate and time consuming (!) the numerical integration.

Special cases can be computed directly from the explicit form of the pdf or the cdf. Quantlets `pdfcauch` and `pdflevy` calculate values of the probability density functions, whereas quantlets `cdfcauch` and `cdflevy` calculate values of the cumulative distribution functions for the Cauchy and Levy distributions, respectively. \mathbf{x} is the input array; `sigma` and `mu` are the scale and location parameters of these distributions.

1.2.2 Simulation of α -stable variables

The complexity of the problem of simulating sequences of α -stable random variables results from the fact that there are no analytic expressions for the inverse F^{-1} of the cumulative distribution function. The first breakthrough was made by Kanter (1975), who gave a direct method for simulating $S_\alpha(1, 1, 0)$ random variables, for $\alpha < 1$. It turned out that this method could be easily adapted to the general case. Chambers, Mallows and Stuck (1976) were the first to give the formulas.

The algorithm for constructing a random variable $X \sim S_\alpha(1, \beta, 0)$, in representation (1.1), is the following (Weron, 1996):

- generate a random variable V uniformly distributed on $(-\frac{\pi}{2}, \frac{\pi}{2})$ and an independent exponential random variable W with mean 1;
- for $\alpha \neq 1$ compute:

$$X = S_{\alpha,\beta} \times \frac{\sin\{\alpha(V + B_{\alpha,\beta})\}}{\{\cos(V)\}^{1/\alpha}} \times \left[\frac{\cos\{V - \alpha(V + B_{\alpha,\beta})\}}{W} \right]^{(1-\alpha)/\alpha}, \quad (1.3)$$

where

$$B_{\alpha,\beta} = \frac{\arctan(\beta \tan \frac{\pi\alpha}{2})}{\alpha},$$

$$S_{\alpha,\beta} = \left\{ 1 + \beta^2 \tan^2 \left(\frac{\pi\alpha}{2} \right) \right\}^{1/(2\alpha)};$$

- for $\alpha = 1$ compute:

$$X = \frac{2}{\pi} \left\{ \left(\frac{\pi}{2} + \beta V \right) \tan V - \beta \log \left(\frac{\frac{\pi}{2} W \cos V}{\frac{\pi}{2} + \beta V} \right) \right\}. \quad (1.4)$$

Given the formulas for simulation of a standard α -stable random variable, we can easily simulate a stable random variable for all admissible values of the parameters α , σ , β and μ using the following property: if $X \sim S_\alpha(1, \beta, 0)$ then

$$Y = \begin{cases} \sigma X + \mu, & \alpha \neq 1, \\ \sigma X + \frac{2}{\pi} \beta \sigma \log \sigma + \mu, & \alpha = 1, \end{cases} \quad (1.5)$$

is $S_\alpha(\sigma, \beta, \mu)$. Although many other approaches have been presented in the literature, this method is regarded as the fastest and the most accurate.

Quantlets `rndstab` and `rndsstab` use formulas (1.3)-(1.5) and provide pseudo random variables of stable and symmetric stable distributions, respectively. Parameters `alpha` and `sigma` in both quantlets and `beta` and `mu` in the first one determine the parameters of the stable distribution.

1.2.3 Tail behavior

Levy (1925) has shown that when $\alpha < 2$ the tails of α -stable distributions are asymptotically equivalent to a Pareto law. Namely, if $X \sim S_{\alpha < 2}(1, \beta, 0)$ then as $x \rightarrow \infty$:

$$P(X > x) = 1 - F(x) \rightarrow C_\alpha(1 + \beta)x^{-\alpha}, \quad (1.6)$$

$$P(X < -x) = F(-x) \rightarrow C_\alpha(1 - \beta)x^{-\alpha},$$

where

$$C_\alpha = \left(2 \int_0^\infty x^{-\alpha} \sin x dx \right)^{-1} = \frac{1}{\pi} \Gamma(\alpha) \sin \frac{\pi\alpha}{2}.$$

The convergence to a power-law tail varies for different α 's (Mandelbrot, 1997, Chapter 14) and, as can be seen in Figure 1.5, is slower for larger values of the tail index. Moreover, the tails of α -stable distribution functions exhibit a crossover from an approximate power decay with exponent $\alpha > 2$ to the true tail with exponent α . This phenomenon is more visible for large α 's (Weron, 2001).

1.3 Estimation of parameters

The estimation of stable law parameters is in general severely hampered by the lack of known closed-form density functions for all but a few members of the stable family. Most of the conventional methods in mathematical statistics, including the maximum likelihood estimation method, cannot be used directly in this case, since these methods depend on an explicit form for the density. However, there are numerical methods that have been found useful in practice and are discussed in this section.

All presented methods work quite well assuming that the sample under consideration is indeed α -stable. However, if the data comes from a different distribution, these procedures may mislead more than the Hill and direct tail estimation methods. Since there are no formal tests for assessing the α -stability of a data set we suggest to first apply the "visual inspection" or non-parametric tests to see whether the empirical densities resemble those of α -stable laws.

Given a sample x_1, \dots, x_n from $S_\alpha(\sigma, \beta, \mu)$, in what follows, we will provide estimates $\hat{\alpha}$, $\hat{\sigma}$, $\hat{\beta}$ and $\hat{\mu}$ of α , σ , β and μ , respectively.

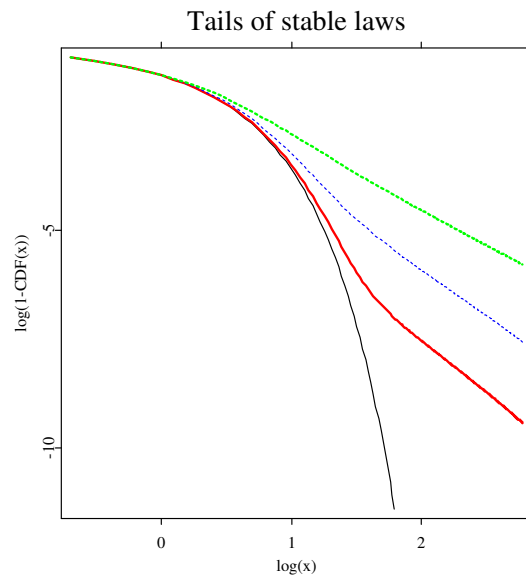


Figure 1.5: Right tails of symmetric α -stable distribution functions for $\alpha = 2$ (thin black), 1.95 (red), 1.8 (thin, dashed blue) and 1.5 (dashed green) on a double logarithmic paper. For $\alpha < 2$ the tails form straight lines with slope $-\alpha$.

■ STFstab05.xpl

1.3.1 Tail exponent estimation

The simplest and most straightforward method of estimating the tail index is to plot the right tail of the (empirical) cumulative distribution function (i.e. $1 - F(x)$) on a double logarithmic paper. The slope of the linear regression for large values of x yields the estimate of the tail index α , through the relation $\alpha = -\text{slope}$.

This method is very sensitive to the sample size and the choice of the number of observations used in the regression. Moreover, the slope around -3.7 may indicate a non- α -stable power-law decay in the tails or the contrary – an α -stable

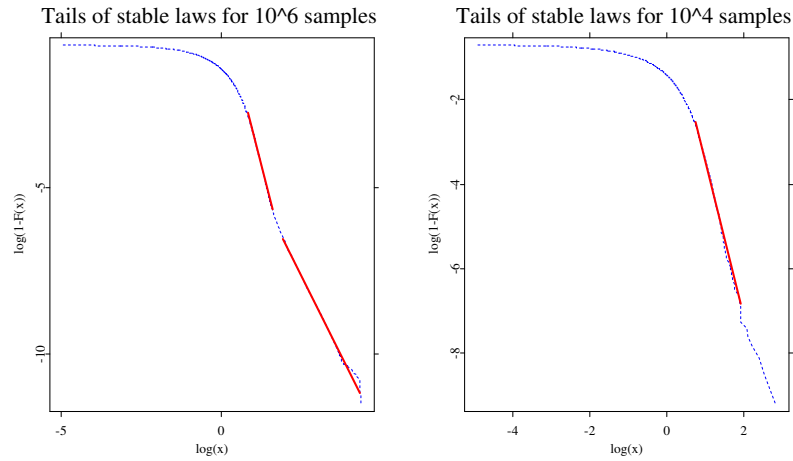


Figure 1.6: A double logarithmic plot of the right tail of an empirical symmetric 1.9-stable distribution function for sample size $N = 10^6$ (left panel) and $N = 10^4$ (right panel). Thick red lines represent the linear regression fit. Even the far tail estimate $\hat{\alpha} = 1.9309$ is above the true value of α . For the smaller sample, the obtained tail index estimate ($\hat{\alpha} = 3.7320$) is close to the initial power-law like decay of the larger sample ($\hat{\alpha} = 3.7881$).

■ STFstab06.xpl

distribution with $\alpha \approx 1.9$. To illustrate this run quantlet `STFstab06`. First simulate (using equation (1.3) and quantlet `rndsstab`) samples of size $N = 10^4$ and 10^6 of standard symmetric ($\beta = \mu = 0$, $\sigma = 1$) α -stable distributed variables with $\alpha = 1.9$. Next, plot the right tails of the empirical distribution functions on a double logarithmic paper, see Figure 1.6.

The true tail behavior (1.6) is observed only for very large (also for very small, i.e. the negative tail) observations, after a crossover from a temporary power-like decay. Moreover, the obtained estimates still have a slight positive bias, which suggests that perhaps even larger samples than 10^6 observations should be used. In Figure 1.6 we used only the upper 0.15% of the records to estimate

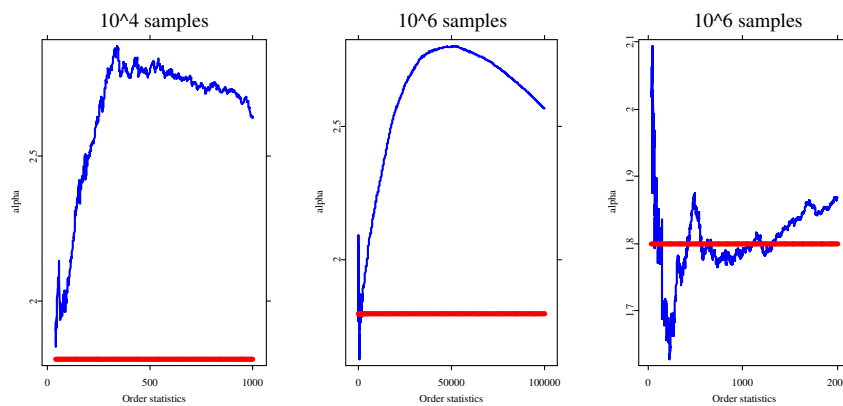


Figure 1.7: Plots of the Hill statistics $\hat{\alpha}_{n,k}$ vs. the maximum order statistic k for 1.8-stable samples of size $N = 10^4$ (left panel) and $N = 10^6$ (middle and right panels). Red horizontal lines represent the true value of α . For better exposition, the right panel is a magnification of the middle panel for small k . A close estimate is obtained only for $k = 500, \dots, 1300$ (i.e. for $k < 0.13\%$ of sample size).

STFstab07.xpl

the true tail exponent. In general, the choice of the observations used in the regression is subjective and can yield large estimation errors, a fact which is often neglected in the literature.

A well known method for estimating the tail index that does not assume a parametric form for the entire distribution function, but focuses only on the tail behavior was proposed by Hill (1975). The Hill estimator is used to estimate the tail index α , when the upper (or lower) tail of the distribution is of the form: $1 - F(x) = Cx^{-\alpha}$. Like the log-log regression method, the Hill estimator tends to overestimate the tail exponent of the stable distribution if α is close to two and the sample size is not very large, see Figure 1.7. For a review of the extreme value theory and the Hill estimator see Chapter 13 in Härdle, Klinken, and Müller (2000) or Embrechts, Klüppelberg and Mikosch (1997).

These examples clearly illustrate that the true tail behavior of α -stable laws is visible only for extremely large data sets. In practice, this means that in order to estimate α we must use high-frequency asset returns and restrict ourselves to the most "outlying" observations. Otherwise, inference of the tail index may be strongly misleading and rejection of the α -stable regime unfounded.

1.3.2 Sample Quantiles Methods

Let x_f be the f -th population quantile, so that $S_\alpha(\sigma, \beta, \mu)(x_f) = f$. Let \hat{x}_f be the corresponding sample quantile, i.e. \hat{x}_f satisfies $F_n(\hat{x}_f) = f$.

McCulloch (1986) analyzed stable law quantiles and provided consistent estimators of all four stable parameters, however, with the restriction $\alpha \geq 0.6$. Define

$$v_\alpha = \frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}, \quad (1.7)$$

which is independent of both σ and μ . Let \hat{v}_α be the corresponding sample value. It is a consistent estimator of v_α . Now, define

$$v_\beta = \frac{x_{0.95} + x_{0.05} - 2x_{0.50}}{x_{0.95} - x_{0.05}}, \quad (1.8)$$

and let \hat{v}_β be the corresponding sample value. v_β is also independent of both σ and μ . As a function of α and β it is strictly increasing in β for each α . The statistic \hat{v}_β is a consistent estimator of v_β .

Statistics v_α and v_β are functions of α and β . This relationship may be inverted and the parameters α and β may be viewed as functions of v_α and v_β

$$\alpha = \psi_1(v_\alpha, v_\beta), \quad \beta = \psi_2(v_\alpha, v_\beta). \quad (1.9)$$

Substituting v_α and v_β by their sample values and applying linear interpolation between values found in tables provided by McCulloch (1986) yields estimators $\hat{\alpha}$ and $\hat{\beta}$.

Scale and location parameters, σ and μ , can be estimated in a similar way. However, due to the discontinuity of the characteristic function for $\alpha = 1$ and $\beta \neq 0$ in representation (1.1), this procedure is much more complicated.

We refer the interested reader to the original work of McCulloch (1986). The quantlet `stabcull` returns estimates of stable distribution parameters from sample x using McCulloch's method.

1.3.3 Sample Characteristic Function Methods

Given an i.i.d. random sample x_1, \dots, x_n of size n , define the *sample characteristic function* by

$$\hat{\phi}(t) = \frac{1}{n} \sum_{j=1}^n e^{itx_j}. \quad (1.10)$$

Since $|\hat{\phi}(t)|$ is bounded by unity all moments of $\hat{\phi}(t)$ are finite and, for any fixed t , it is the sample average of i.i.d. random variables $\exp(itx_j)$. Hence, by the law of large numbers, $\hat{\phi}(t)$ is a consistent estimator of the characteristic function $\phi(t)$.

Press (1972) proposed a simple estimation method, called the method of moments, based on transformations of the characteristic function. From (1.1) we have for all α

$$|\phi(t)| = \exp(-\sigma^\alpha |t|^\alpha). \quad (1.11)$$

Hence, $-\log |\phi(t)| = \sigma^\alpha |t|^\alpha$. Now, assuming $\alpha \neq 1$, choose two nonzero values of t , say $t_1 \neq t_2$. Then for $k = 1, 2$ we have

$$-\log |\phi(t_k)| = \sigma^\alpha |t_k|^\alpha. \quad (1.12)$$

Solving these two equations for α and σ , and substituting $\hat{\phi}(t)$ for $\phi(t)$ yields

$$\hat{\alpha} = \frac{\log \frac{\log |\hat{\phi}(t_1)|}{\log |\hat{\phi}(t_2)|}}{\log \left| \frac{t_1}{t_2} \right|}, \quad (1.13)$$

and

$$\log \hat{\sigma} = \frac{\log |t_1| \log(-\log |\hat{\phi}(t_2)|) - \log |t_2| \log(-\log |\hat{\phi}(t_1)|)}{\log \left| \frac{t_1}{t_2} \right|}. \quad (1.14)$$

In order to estimate β and μ we have to apply a similar trick to $\Im\{\log \phi(t)\}$. The estimators are consistent since they are based upon estimators of $\phi(t)$, $\Im\{\phi(t)\}$ and $\Re\{\phi(t)\}$, which are known to be consistent. However, convergence to the population values depends on the choice of t_1, \dots, t_4 . The optimal selection of these values is problematic and still is an open question.

The quantlet `stabmom` returns estimates of stable distribution parameters from sample `x` using the method of moments. It uses a selection of points suggested by Koutrouvelis (1980): $t_1 = 0.2$, $t_2 = 0.8$, $t_3 = 0.1$, and $t_4 = 0.4$.

Parameter estimates can be also obtained by minimizing some function of the difference between the theoretical and sample characteristic functions. Koutrouvelis (1980) presented a regression-type method which starts with an initial estimate of the parameters and proceeds iteratively until some prespecified convergence criterion is satisfied. Each iteration consists of two weighted regression runs. The number of points to be used in these regressions depends on the sample size and starting values of α . Typically no more than two or three iterations are needed. The speed of the convergence, however, depends on the initial estimates and the convergence criterion.

The regression method is based on the following observations concerning the characteristic function $\phi(t)$. First, from (1.1) we can easily derive

$$\log(-\log |\phi(t)|^2) = \log(2\sigma^\alpha) + \alpha \log |t|. \quad (1.15)$$

The real and imaginary parts of $\phi(t)$ are for $\alpha \neq 1$ given by

$$\Re\{\phi(t)\} = \exp(-|\sigma t|^\alpha) \cos \left[\mu t + |\sigma t|^\alpha \beta \text{sign}(t) \tan \frac{\pi\alpha}{2} \right],$$

and

$$\Im\{\phi(t)\} = \exp(-|\sigma t|^\alpha) \sin \left[\mu t + |\sigma t|^\alpha \beta \text{sign}(t) \tan \frac{\pi\alpha}{2} \right].$$

The last two equations lead, apart from considerations of principal values, to

$$\arctan\left(\frac{\Im\{\phi(t)\}}{\Re\{\phi(t)\}}\right) = \mu t + \beta\sigma^\alpha \tan\frac{\pi\alpha}{2} \text{sign}(t)|t|^\alpha. \quad (1.16)$$

Equation (1.15) depends only on α and σ and suggests that we estimate these parameters by regressing $y = \log(-\log|\phi_n(t)|^2)$ on $w = \log|t|$ in the model

$$y_k = m + \alpha w_k + \epsilon_k, \quad k = 1, 2, \dots, K, \quad (1.17)$$

where t_k is an appropriate set of real numbers, $m = \log(2\sigma^\alpha)$, and ϵ_k denotes an error term. Koutrouvelis (1980) proposed to use $t_k = \frac{\pi k}{25}$, $k = 1, 2, \dots, K$; with K ranging between 9 and 134 for different estimates of α and sample sizes.

Once $\hat{\alpha}$ and $\hat{\sigma}$ have been obtained and α and σ have been fixed at these values, estimates of β and μ can be obtained using (1.16). Next, the regressions are repeated with $\hat{\alpha}$, $\hat{\sigma}$, $\hat{\beta}$ and $\hat{\mu}$ as the initial parameters. The iterations continue until a prespecified convergence criterion is satisfied.

Kogon and Williams (1998) eliminated this iteration procedure and simplified the regression method. For initial estimation they applied McCulloch's (1986) method, worked with the continuous representation (1.2) of the characteristic function instead of the classical one (1.1) and used a fixed set of only 10 equally spaced frequency points t_k . In terms of computational speed their method compares favorably to the original method of Koutrouvelis (1980). It has a significantly better performance near $\alpha = 1$ and $\beta \neq 0$ due to the elimination of discontinuity of the characteristic function. However, it returns slightly worse results for very small α .

The quantlet `stabreg` fits a stable distribution to sample \mathbf{x} and returns parameter estimates. The string `method` determines the method used: `method="k"` denotes the Koutrouvelis (1980) method with McCulloch's (1986) initial parameter estimates (default), `method="km"` denotes the Koutrouvelis (1980) method with initial parameter estimates obtained from the method of moments, and `method="kw"` denotes the Kogon and Williams (1998) method. The last two optional parameters are responsible for computation accuracy: `epsilon` (default `epsilon=0.00001`) specifies the convergence criterion, whereas `maxit` (default `maxit=5`) denotes the maximum number of iterations for both variants of the Koutrouvelis (1980) method.

1.4 Financial applications of α -stable laws

Distributional assumptions for financial processes have important theoretical implications, given that financial decisions are commonly based on expected returns and risk of alternative investment opportunities. Hence, solutions to such problems like portfolio selection, option pricing, and risk management depend crucially on distributional specifications.

Many techniques in modern finance rely heavily on the assumption that the random variables under investigation follow a Gaussian distribution. However, time series observed in finance – but also in other econometric applications – often deviate from the Gaussian model, in that their marginal distributions are heavy-tailed and, possibly, asymmetric. In such situations, the appropriateness of the commonly adopted normal assumption is highly questionable.


It is often argued that financial asset returns are the cumulative outcome of a vast number of pieces of information and individual decisions arriving almost continuously in time. Hence, in the presence of heavy-tails it is natural to assume that they are approximately governed by a stable non-Gaussian distribution. Other leptokurtic distributions, including Student's t , Weibull and hyperbolic, lack the attractive central limit property.

Stable distributions have been successfully fit to stock returns, excess bond returns, foreign exchange rates, commodity price returns and real estate returns (McCulloch, 1996; Rachev and Mittnik, 2000). In recent years, however, several studies have found, what appears to be strong evidence against the stable model (for a review see: McCulloch, 1997; Weron, 2001). These studies have estimated the tail exponent directly from the tail observations and commonly have found α that appears to be significantly greater than 2, well outside the stable domain. Recall, however, that in Section 1.3.1 we have shown that estimating α only from the tail observations may be strongly misleading and for samples of typical size the rejection of the α -stable regime unfounded. Let us see ourselves how well the stable law describes financial asset returns.

The quantlet `STFstab08` fits the stable and Gaussian laws to the USD/GBP exchange rate returns from the period January 2, 1990 – November 8, 2000 using the regression method of Koutrouvelis (1980). Next, it compares both fits through Anderson-Darling (see Stephens, 1974) and Kolmogorov test statistics. The former may be treated as a weighted Kolmogorov statistics which puts more weight to the differences in the tails of the distributions. The obtained results, presented in Table 1.1, clearly show that the 1.71-stable law

Table 1.1: α -stable and Gaussian fits to the USD/GBP exchange rate returns from the period January 2, 1990 – November 8, 2000.

Parameters	α	σ	β	μ
α -stable fit	1.7165	0.0033	-0.0986	0.0000
Gaussian fit		0.0059		0.0000
Test values	Anderson-Darling		Kolmogorov	
α -stable fit	2.1209		0.9181	
Gaussian fit	20.5433		3.1816	

 STFstab08.xpl
Table 1.2: α -stable and Gaussian fits to the Dow Jones Industrial Average (DJIA) index from the period July 6, 1984 – May 17, 1996.

Parameters	α	σ	β	μ
α -stable fit	1.6723	0.0048	0.0999	0.0009
Gaussian fit		0.0102		0.0006
Test values	Anderson-Darling		Kolmogorov	
α -stable fit	1.0624		0.8046	
Gaussian fit	+INF		4.9778	

 STFstab09.xpl

offers a much better fit to the data than the Gaussian.

The quantlet `STFstab09` fits both distributions to the Dow Jones Industrial Average (DJIA) index from the period July 6, 1984 – May 17, 1996. Recall, that this period includes the biggest crash in stock market's history – the Black Monday of October 19, 1987. Clearly the 1.67-stable law offers a much better fit to the DJIA returns, see Table 1.2. Its superiority, especially in the tails of the distribution, is even better visible in Figure 1.8.

The quantlet `STFstab10` fits both laws to the Deutsche Aktienindex (DAX) index from the period January 2, 1995 – December 11, 2002. Like for two previous data sets also here the α -stable law offers a much better fit, see Table

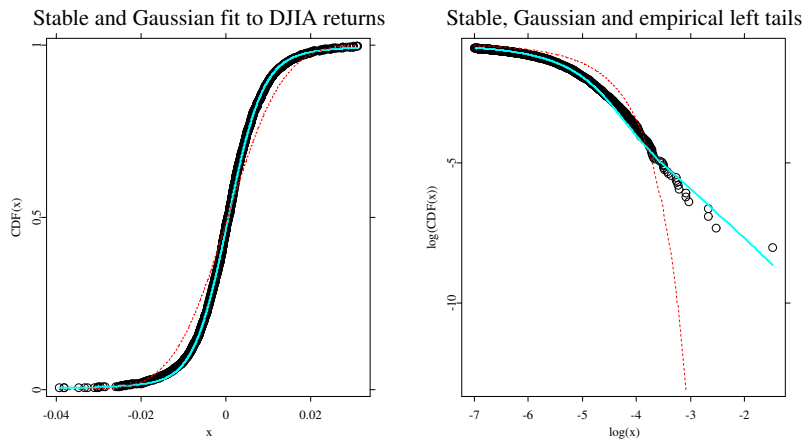


Figure 1.8: 1.67-stable (cyan) and Gaussian (dashed red) fits to the DJIA returns (black circles) empirical cumulative distribution function from the period July 6, 1984 – May 17, 1996. Right panel is a magnification of the left tail fit on a double logarithmic scale clearly showing the superiority of the 1.67-stable law.


STFstab09.xpl

1.3. In all three cases the Gaussian distribution has to be rejected based on the results of the two applied tests – the p -values are less than 10^{-4} . Unfortunately, the critical test values are not known for the α -stable law. However, the Anderson-Darling test values are at least an order of magnitude smaller and the Kolmogorov test values are a few times smaller than for the Gaussian distribution suggesting a much better fit, especially in the tails.

Of course, nothing we have said demonstrates that any one of the above data sets really has a stable distribution. Although we have shown in Section 1.3.1 that tail index estimates are not sufficient to reject stability, the converse is also true: an estimate of α significantly less than 2 by no means rules out a non-stable distribution with power-law tails with $\alpha > 2$. Even though it is relatively easy to reject normality (as we have shown above), alternative leptokurtic distributions are very hard to tell apart.

Table 1.3: α -stable and Gaussian fits to the the Deutsche Aktienindex (DAX) index from the period January 2, 1995 – December 11, 2002.

Parameters	α	σ	β	μ
α -stable fit	1.6975	0.0088	-0.3255	-0.0003
Gaussian fit		0.0158		0.0003
Test values	Anderson-Darling		Kolmogorov	
α -stable fit	1.9714		1.1625	
Gaussian fit	16.5865		2.8499	

 STFstab10.xpl

Yet, the central limit property of stable laws, together with good description of extreme events may justify their application to diverse problems in finance including portfolio optimization, option pricing and – most noticeably – Value-at-Risk type calculations, where the estimation of low quantiles in portfolio return distributions is crucial (Khindarova, Rachev, and Schwartz, 2001).

Bibliography

- Chambers, J. M., Mallows, C. L. and Stuck, B. W. (1976). A Method for Simulating Stable Random Variables, *Journal of the American Statistical Association* **71**: 340–344.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer.
- Fofack, H. and Nolan, J. P. (1999). Tail Behavior, Modes and Other Characteristics of Stable Distributions, *Extremes* **2** (1999): 39–58.
- Härdle, W., Klinke, S., and Müller, M. (2000). *XploRe Learning Guide*, Springer.
- Hill, B. M. (1975). A Simple General Approach to Inference About the Tail of a Distribution, *Annals of Statistics* **3**: 1163–1174.
- Janicki, A. and Weron, A. (1994). *Simulation and Chaotic Behavior of α -Stable Stochastic Processes*, Marcel Dekker.
- Kanter, M. (1975). Stable Densities Under Change of Scale and Total Variation Inequalities, *Annals of Probability* **3**: 697–707.
- Khindarova, I., Rachev, S. and Schwartz, E. (2001). Stable Modeling of Value at Risk, *Mathematical and Computer Modelling* **34**: 1223–1259.
- Koutrouvelis, I. A. (1980). Regression-Type Estimation of the Parameters of Stable Laws, *Journal of the American Statistical Association* **75**: 918–928.
- Kogon, S. M. and Williams, D. B. (1998). Characteristic function based estimation of stable parameters, in R. Adler, R. Feldman, M. Taqqu (eds.), *A Practical Guide to Heavy Tails*, Birkhauser, pp. 311–335.
- Levy, P. (1925). *Calcul des Probabilites*, Gauthier Villars.

- Mandelbrot, B. B. (1997). *Fractals and Scaling in Finance*, Springer.
- McCulloch, J. H. (1986). Simple Consistent Estimators of Stable Distribution Parameters, *Communications in Statistics – Simulations* **15**: 1109–1136.
- McCulloch, J. H. (1996). Financial Applications of Stable Distributions, in G. S. Maddala, C. R. Rao (eds.), *Handbook of Statistics, Vol. 14*, Elsevier, pp. 393–425.
- McCulloch, J. H. (1997). Measuring Tail Thickness to Estimate the Stable Index α : A Critique, *Journal of Business & Economic Statistics* **15**: 74–81.
- Nolan, J. P. (1997). Numerical Calculation of Stable Densities and Distribution Functions, *Communications in Statistics – Stochastic Models* **13**: 759–774.
- Press, S. J. (1972). Estimation in Univariate and Multivariate Stable Distribution, *Journal of the American Statistical Association* **67**: 842–846.
- Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance*, Wiley.
- Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*, Chapman & Hall.
- Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association* **69**: 730–737.
- Uchaikin, V. V. and Zolotarev, V. M. (1999). *Chance and Stability: Stable Distributions and their Applications*, VSP.
- Weron, R. (1996). On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables, *Statistics and Probability Letters* **28**: 165–171. See also R. Weron, *Correction to: On the Chambers-Mallows-Stuck Method for Simulating Skewed Stable Random Variables*, Research Report HSC/96/1, Wrocław University of Technology, 1996, <http://www.im.pwr.wroc.pl/~hugo/Publications.html>.
- Weron, R. (2001). Levy–Stable Distributions Revisited: Tail Index > 2 Does Not Exclude the Levy–Stable Regime, *International Journal of Modern Physics C* **12**: 209–223.
- Zolotarev, V. M. (1986). *One-Dimensional Stable Distributions*, American Mathematical Society.

2 Tail dependence

Rafael Schmidt

Tail dependence describes the amount of dependence in the lower-left-quadrant tail or upper-right-quadrant tail of a bivariate distribution. Multivariate distributions possessing tail dependence are able to incorporate dependencies of extremal events. According to Hauksson *et al.* (2001), Resnick (2002), and Embrechts *et al.* (2001) tail dependence plays an important role in extreme value theory, finance, and insurance models. In particular in credit-portfolio modelling tail-dependent distributions are of special practical interest, as dependencies of large credit-default events can be modelled.

Here we introduce the tail-dependence concept and characterizing tail dependence by the so-called tail-dependence coefficient and we embed this concept into the general framework of copulae. In Section ??, the tail-dependence coefficient is calculated for several Archimedean copulae, elliptically contoured distributions, and other copulae. Section ?? proposes two non-parametric estimators for the tail-dependence coefficient based on the empirical copula. Further we provide a suitable estimator for the tail-dependence coefficient within the class of elliptically-contoured distributions. The last section presents empirical results comparing the latter estimators.

2.1 Tail dependence and copulae

Tail-dependence definitions for multivariate random vectors are mostly related to their bivariate marginal distribution functions. Loosely speaking, tail dependence describes the limiting proportion of exceeding one margin over a certain threshold given that the other margin has already exceeded that threshold. The following approach from Joe (1997) represents one of many possible definitions of tail dependence.

Let $X = (X_1, X_2)^\top$ be a 2-dimensional random vector. We say that X is

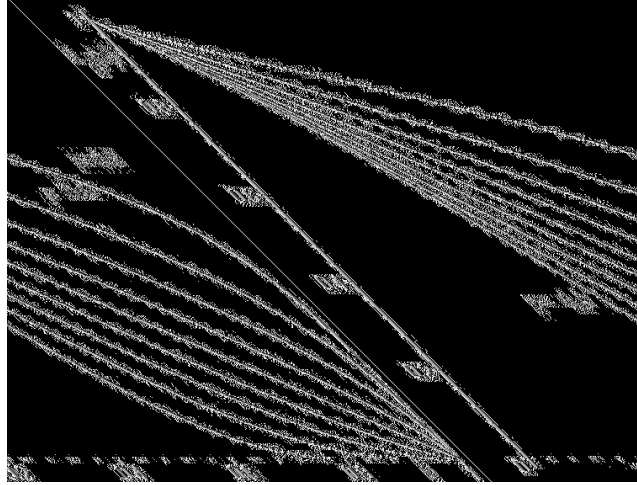


Figure 2.1: $\lambda_U(v) = \mathbb{P}\{X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v)\}$ for a bivariate normal distribution with $\rho = -0.8, -0.6, \dots, 0.6, 0.8$. Note that $\lambda_U = 0$ for all $\rho \in (-1, 1)$.

(bivariate) *upper tail-dependent* if

$$\lambda_U := \lim_{v \rightarrow 1^-} \lambda_U(v) = \lim_{v \rightarrow 1^-} \mathbb{P}\{X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v) > 0\}, \quad (2.1)$$

in case the limit exists. F_1^{-1}, F_2^{-1} denote the generalized inverse distribution functions of X_1, X_2 . Consequently, we say $X = (X_1, X_2)^\top$ is *upper tail-independent* if λ_U equals 0. Further, we call λ_U the upper tail-dependence coefficient (TDC). See Schmidt and Stadtmüller (2002) for a generalization to multivariate tail dependence. Similarly, we define the lower tail-dependence coefficient by

$$\lambda_L := \lim_{v \rightarrow 0^+} \mathbb{P}\{X_1 \leq F_1^{-1}(v) \mid X_2 \leq F_2^{-1}(v)\}. \quad (2.2)$$

We embed the tail-dependence concept within the copula theory. An n -dimensional distribution function $C : [0, 1]^n \rightarrow [0, 1]$ is called a copula if it has one-dimensional margins which are uniformly distributed on the interval $[0, 1]$. Copulae are functions that join or "couple" an n -dimensional distribution function F to its corresponding one-dimensional marginal-distribution functions

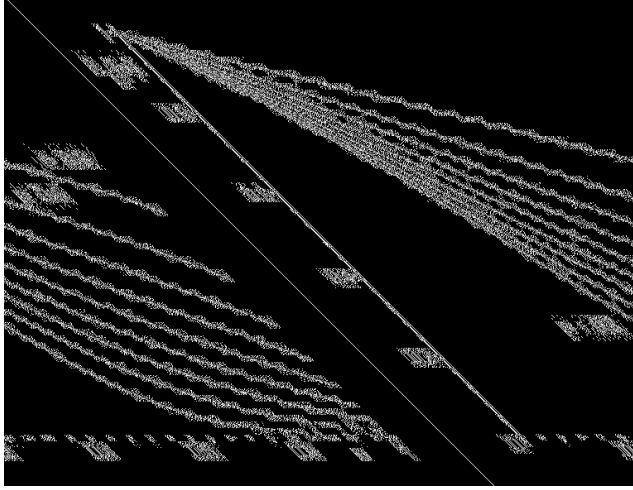


Figure 2.2: $\lambda_U(v) = \mathbb{P}\{X_1 > F_1^{-1}(v) \mid X_2 > F_2^{-1}(v)\}$ for a bivariate t -distribution with $\rho = -0.8, -0.6, \dots, 0.6, 0.8$. The exact formula for $\lambda_U > 0$ is given in Section ??.

F_i , $i = 1, \dots, n$, in the following way:

$$F(x_1, \dots, x_n) = C\{F_1(x_1), \dots, F_n(x_n)\}.$$

We refer the reader to the monographs of Nelsen (1999) or Joe (1997) for more information on copulae.

The following representation shows that tail dependence is a copula property. Thus many copula features transfer to the tail-dependence coefficient, for example the invariance under strictly increasing transformations of the margins.

If X is a continuous bivariate random vector, then

$$\lambda_U = \lim_{v \rightarrow 1^-} \frac{1 - 2v + C(v, v)}{1 - v}, \quad (2.3)$$

where C denotes the copula of X . Analogously, $\lambda_L = \lim_{v \rightarrow 0^+} \frac{C(v, v)}{v}$ holds for the lower tail-dependence coefficient.

2.2 Calculating the tail-dependence coefficient

2.2.1 Archimedean copulae

Referring to Härdle, Kleinow and Stahl (2002) Archimedean copulae are implemented and investigated within the Value at Risk framework. Therefore and because Archimedean copulae form an important class of copulae which are easy to construct and have nice properties, we explore the tail-dependence coefficient for this family of copulae. A bivariate Archimedean copula has the form $C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v))$ for some continuous, strictly decreasing, and convex function $\phi : [0, 1] \rightarrow [0, \infty]$ such that $\phi(1) = 0$ and the pseudo-inverse function $\phi^{[-1]}$ is defined by

$$\phi^{[-1]}(t) = \begin{cases} \phi^{-1}(t), & 0 \leq t \leq \phi(0), \\ 0, & \phi(0) < t \leq \infty. \end{cases}$$

We call ϕ *strict* if $\phi(0) = \infty$. In that case $\phi^{[-1]} = \phi^{-1}$. Regarding the existence of tail dependence for Archimedean copulae we can show that:

1. Upper tail-dependence implies $\phi'(1) = 0$
and $\lambda_U = 2 - \lim_{v \rightarrow 1^-} (\phi^{[-1]} \circ 2\phi)'(v)$,
2. $\phi'(1) < 0$ implies upper tail-independence,
3. $\phi'(0) > -\infty$ or a non-strict ϕ implies lower tail-independence,
4. Lower tail-dependence implies $\phi'(0) = -\infty$, $\lambda_L = \lim_{v \rightarrow 0^+} (\phi^{-1} \circ 2\phi)'(v)$,
and a strict ϕ .

Table 2.1 lists various Archimedean copulae in the same ordering as in Table 2.1 in Härdle, Kleinow and Stahl (2002) or in Nelsen (1999) and the corresponding upper and lower TDC.

The quantlet `TailCoeffCopula` derives the TDC for a large number of bivariate copula-functions. The inputs of this quantlet are the type of copula and the copula parameters. The type of copula is specified by an integer, between 1 and 34, listed in the following tables. For instance `copula=1` - Pareto-Clayton, `copula=4` - Gumbel-Hougaard, and `copula=5` - Frank. The result of the application is assigned to the vector `(lTDC, uTDC)` and contains the lower and upper TDC, respectively. We refer the reader to the `VARCopula` quantlet for related

copula calculations.

```
(lTDC,uTDC) = TailCoeffCopula (copula,parameters)
    Calculates the lower and upper tail-dependence coefficient for various copulae
```

2.2.2 Elliptically contoured distributions

In this section, we calculate the tail-dependence coefficient for elliptically contoured distributions (briefly: elliptical distributions). Well-known elliptical distributions are the multivariate normal distribution, the multivariate t -distribution, the multivariate logistic distribution, and the multivariate symmetric generalized-hyperbolic distribution.

Elliptical distributions are defined as follows: Let X be an n -dimensional random vector and $\Sigma \in \mathbb{R}^{n \times n}$ be a symmetric positive semi-definite matrix. If $X - \mu$, for some $\mu \in \mathbb{R}^n$, possesses a characteristic function of the form $\phi_{X-\mu}(t) = \Phi(t^\top \Sigma t)$ for some function $\Phi : \mathbb{R}_0^+ \rightarrow \mathbb{R}$, then X is said to be elliptically distributed with parameters μ (location), Σ (dispersion), and Φ . Let $E_n(\mu, \Sigma, \Phi)$ denote the class of elliptically contoured distributions with the latter parameters. We call Φ the *characteristic generator*.

The density function, if existent, of an elliptically contoured distribution has the following form:

$$f(x) = |\Sigma|^{-1/2} g(x - \mu)^\top \Sigma^{-1} (x - \mu), \quad x \in \mathbb{R}^n, \quad (2.4)$$

for some function $g : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$, which we call the *density generator*. Observe that the name "elliptically contoured" distribution is related to the elliptical contours of the latter density. For a more detailed treatment of elliptical distributions see Fang, Kotz, and Ng (1990), Cambanis, Huang, and Simon (1981).

Bingham and Kiesel (2002) and Bingham, Kiesel, and Schmidt (2002) propose a semi-parametric approach for financial modelling by estimating the parametric components (μ, Σ) and non-parametric components (density generator g) of an elliptical distribution. Regarding the tail-dependence coefficient, it suffices to consider the tail behavior of the density generator.

Table 2.1: Tail-dependence coefficient (TDC) for various Archimedean copulae. Numbers correspond to table 4.1 in Nelsen (1999), p. 94.

Number & Type	$C(u, v)$	$\phi_\theta(t)$	$\theta \in$	upper-TDC	lower-TDC
(1) Pareto	$\max([u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0)$	$(t^{-\theta} - 1)/\theta$	$[-1, \infty) \setminus \{0\}$	0 for $\theta > 0$	$2^{-1/\theta}$ for $\theta > 0$
(2)	$\max\left(1 - \left[(1-u)^\theta + (1-v)^\theta\right]^{1/\theta}, 0\right)$	$(1-t)^\theta$	$[1, \infty)$	$2 - 2^{1/\theta}$	0
(3) Ali Mikhail-Haq	$\frac{uv}{1 - \theta(1-u)(1-v)}$	$\log \frac{1 - \theta(1-t)}{t}$	$[-1, 1)$	0	0
(4) Gumbel- Hougaard	$\exp\left(-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{1/\theta}\right)$	$(-\log t)^\theta$	$[1, \infty)$	$2 - 2^{1/\theta}$	0
(12)	$\left(1 + \left[(u^{-1} - 1)^\theta + (v^{-1} - 1)^\theta\right]^{1/\theta}\right)^{-1}$	$\left(\frac{1}{t} - 1\right)^\theta$	$[1, \infty)$	$2 - 2^{1/\theta}$	$2^{-1/\theta}$
(14)	$\left(1 + \left[(u^{-1/\theta} - 1)^\theta + (v^{-1/\theta} - 1)^\theta\right]^{1/\theta}\right)^{-\theta}$	$(t^{-1/\theta} - 1)^\theta$	$[1, \infty)$	$2 - 2^{1/\theta}$	$\frac{1}{2}$
(19)	$\theta / \log(e^{\theta/u} + e^{\theta/v} - e^\theta)$	$e^{\theta/t} - e^\theta$	$(0, \infty)$	0	1

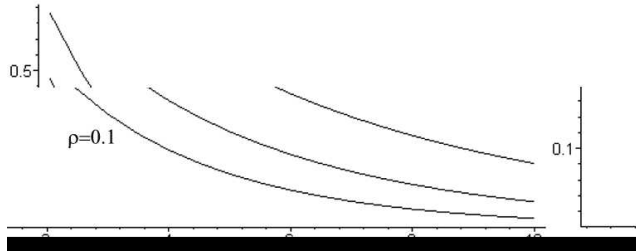


Figure 2.3: Tail-dependence coefficient λ versus α for $\rho = 0.5, 0.3, 0.1$

Schmidt (2002b) shows that bivariate elliptically-contoured distributions are upper and lower tail-dependent if the tail of their density generator is regularly varying. Further a necessary condition for tail dependence is given which is somewhat weaker than regular variation of the latter tail: precisely, the tail must be O-regularly varying.

Although the equivalence of tail dependence and regularly-varying density generator cannot be shown, all density generators of well-known elliptical distributions are given either by a regularly varying tail or a not O-regularly varying tail. This justifies a restriction to the class of elliptical distributions with regularly varying density generator if one wants to incorporate tail dependence. A reference for regular and O-regular variation is Bingham, Goldie, and Teugels (1987).

The following closed-form expression exists (Schmidt, 2002b) for the upper and lower tail-dependence coefficient of an elliptically-contoured random vector $(X_1, X_2)^\top \in E_2(\mu, \Sigma, \Phi)$ with positive-definite matrix Σ , having a regularly

varying density generator g with regular-variation index $-\alpha/2 - 1 < 0$:

$$\lambda := \lambda_U = \lambda_L = \frac{\int_0^{h(\rho)} \frac{u^\alpha}{\sqrt{1-u^2}} du}{\int_0^1 \frac{u^\alpha}{\sqrt{1-u^2}} du}, \quad (2.5)$$

where $\rho := \Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$ and $h(\rho) := \left(1 + \frac{(1-\rho)^2}{1-\rho^2}\right)^{-1/2}$ (see also Figure 2.3). Note that ρ corresponds to the "correlation" coefficient when this exists (Fang, Kotz, and Ng, 1990). Moreover, the upper tail-dependence coefficient λ_U coincides with the lower tail-dependence coefficient λ_L and depends only on the "correlation" coefficient ρ and the regular-variation index α .

Table 2.2 lists various elliptical distributions, the corresponding density generators (here c_n denotes a normalizing constant depending only on the dimension n) and the associated tail index α from which one easily derives the tail-dependence coefficient using formula (2.5).

2.2.3 Other copulae

For many other closed-form copulae one can explicitly derive the tail-dependence coefficient. Table 2.3 lists various well-known copula-functions and the corresponding lower and upper TDC. Most copulae are listed in the monograph of Joe (1997).

Table 2.2: Tail index α for various density generators g of multivariate elliptical distributions. (K_ν denotes the modified Bessel function of the third kind (or Macdonald function))

Number & Type	Density generator g or characteristic generator Φ	Parameters & α	for $n = 2$
(23) Normal	$g(u) = c_n \exp(-u/2)$	-	∞
(24) t	$g(u) = c_n \left(1 + \frac{t}{\theta}\right)^{-(n+\theta)/2}$	$\theta > 0$	θ
(25) Symmetric general. hyperbolic	$g(u) = c_n \frac{K_{\lambda-\frac{n}{2}}(\sqrt{\psi(\chi+u)})}{(\sqrt{\chi+u})^{\frac{n}{2}-\lambda}}$	$\psi, \chi > 0$ $\lambda \in \mathbb{R}$	∞
(26) Symmetric θ -stable	$\Phi(u) = \exp\left[-\left(\frac{1}{2}u\right)^{\theta/2}\right]$	$\theta \in (0, 2]$	θ
(27) logistic	$g(u) = c_n \frac{\exp(-u)}{(1 + \exp(-u))^2}$	-	∞

Table 2.3: Tail-dependence coefficient (TDC) for various copulae. (copulae BBx are given in Joe (1997))

Number & Type	$C(u, v)$		
	Parameters	upper-TDC	lower-TDC
(28) Raftery	$g(x, y; \theta) = x - \frac{1-\theta}{1+\theta} x^{1/(1-\theta)} \left(y^{-\theta/(1-\theta)} - y^{1/(1-\theta)} \right)$	$\theta \in [0, 1]$	$\frac{2\theta}{1+\theta}$
(29) BB1	$\left[1 + \left\{ (u^{-\theta} - 1)^\delta + (v^{-\theta} - 1)^\delta \right\}^{1/\delta} \right]^{-1/\theta}$	$\theta \in (0, \infty)$ $\delta \in [1, \infty)$	$2 - 2^{1/\delta}$ $2^{-1/(\theta\delta)}$
(30) BB4	$\left[u^{-\theta} + v^{-\theta} - 1 - \left\{ (u^{-\theta} - 1)^{-\delta} + (v^{-\theta} - 1)^{-\delta} \right\}^{-1/\delta} \right]^{-1/\theta}$	$\theta \in [0, \infty)$ $\delta \in (0, \infty)$	$2^{-1/\delta}$ $2^{-1/(\delta)} - 1/\theta$
(31) BB7	$1 - \left(1 - \left[\{1 - (1-u)^\theta\}^{-\delta} + \{1 - (1-v)^\theta\}^{-\delta} - 1 \right]^{-1/\delta} \right)^{1/\theta}$	$\theta \in [1, \infty)$ $\delta \in (0, \infty)$	$2 - 2^{1/\theta}$ $2^{-1/\delta}$
(32) BB8	$\frac{1}{\delta} \left(1 - \left[1 - \{1 - (1-\delta)^\theta\}^{-1} \cdot \{1 - (1-\delta u)^\theta\} \{1 - (1-\delta v)^\theta\} \right]^{1/\theta} \right)$	$\theta \in [1, \infty)$ $\delta \in [0, 1]$	2^{-} $2(1-\delta)^{\theta-1}$ 0
(33) BB11	$\theta \min(u, v) + (1-\theta)uv$	$\theta \in [0, 1]$	θ
(34) C_Ω in Junker et al. (2002)	$\beta C_{(\theta, \delta)}^s(u, v) - (1-\beta)C_{(\theta, \delta)}(u, v)$ with Archimedean generator $\phi_{(\theta, \delta)}(t) = \left(-\log \frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)^\delta$ $C_{(\theta, \delta)}^s$ is the corresp. survival copula with param. (θ, δ)	$\theta, \beta \in \mathbb{R} \setminus \{0\}$ $\delta, \delta \geq 1$ $\beta \in [0, 1]$	$\beta(2 - 2^{1/\delta})$ $(1-\beta) \cdot$ $(2 - 2^{1/\delta})^{\theta-1}$

2.3 Estimating the tail-dependence coefficient

Suppose $X, X^{(1)}, \dots, X^{(m)}$ are i.i.d. bivariate random vectors with distribution function F and copula C . We assume continuous marginal distribution functions F_i , $i = 1, 2$. Techniques testing for tail dependence or tail independence are given for example in Ledford and Tawn (1996).

We propose the following two non-parametric estimators for the lower and upper tail-dependence coefficient λ_U, λ_L , based on the copula representation (2.3). Let C_m be the empirical copula defined by

$$C_m(u, v) = F_m, F_{1m}^{-1}(u), F_{2m}^{-1}(v) \quad (2.6)$$

with F_m and F_{im} denoting the empirical distribution functions corresponding to F, F_i , $i = 1, 2$. Let $R_{m1}^{(j)}$ and $R_{m2}^{(j)}$ be the rank of $X_1^{(j)}$ and $X_2^{(j)}$, $j = 1, \dots, m$, respectively. The first estimators are based on formulas (2.1) and (2.2):

$$\hat{\lambda}_{U,m}^{(1)} = \frac{m}{k} C_m \left\{ \left(1 - \frac{k}{m}, 1\right) \times \left(1 - \frac{k}{m}, 1\right) \right\} = \frac{1}{k} \sum_{j=1}^m \mathbf{1}_{\{R_{m1}^{(j)} > m-k, R_{m2}^{(j)} > m-k\}} \quad (2.7)$$

and

$$\hat{\lambda}_{L,m}^{(1)} = \frac{m}{k} C_m \left(\frac{k}{m}, \frac{k}{m} \right) = \frac{1}{k} \sum_{j=1}^m \mathbf{1}_{\{R_{m1}^{(j)} \leq k, R_{m2}^{(j)} \leq k\}}, \quad (2.8)$$

where $k = k(m) \rightarrow \infty$ and $k/m \rightarrow 0$ as $m \rightarrow \infty$, and the first expression in (2.7) has to be understood as the empirical copula-measure of the interval $(1 - k/m, 1] \times (1 - k/m, 1]$. The second estimators are based on techniques from extreme-value theory. We refer the reader to Schmidt and Stadtmüller (2002) for more details.

$$\begin{aligned} \hat{\lambda}_{U,m}^{(2)} &= 2 - \frac{m}{k} \left\{ 1 - C_m \left(1 - \frac{k}{m}, 1 - \frac{k}{m} \right) \right\} \\ &= 2 - \frac{1}{k} \sum_{j=1}^m \mathbf{1}_{\{R_{m1}^{(j)} > m-k \text{ or } R_{m2}^{(j)} > m-k\}}, \end{aligned} \quad (2.9)$$

with $k = k(m) \rightarrow \infty$ and $k/m \rightarrow 0$ as $m \rightarrow \infty$. $\hat{\lambda}_{L,m}^{(2)}$ is defined similarly. The optimal choice of k is related to the usual variance-bias problem, which we investigate in a forthcoming work. Strong consistency and asymptotic normality are addressed in Schmidt and Stadtmüller (2002). See also Schmidt (2002a) for related results within the class of elliptically-contoured distributions and copulae.

The quantlet `TailCoeffEstimation` estimates the upper and lower TDC based on formulas (2.7) and (2.8). The inputs of this quantlet are the two-dimensional data-set, the threshold k , and whether the lower or upper TDC is calculated. The result of the application is assigned to the variable `TDC` and contains the estimated lower or upper TDC, respectively.

```
TDC = TailCoeffEstimation(data, threshold, upper)
Estimates the upper and lower tail-dependence coefficient based
on formulas (2.7) and (2.8).
```

Now we focus on an elliptically-contoured bivariate random vector X . In the presence of tail dependence, arguments from Section ?? justify considering only elliptical distributions having a regularly varying density generator with tail index α . According to Schmidt (2002b), this implies that the distribution function of $\|X\|_2$ has also a regularly varying tail with index α ($\|\cdot\|_2$ denotes the Euclidean norm). Formula (2.5) shows that the upper and lower tail-dependence coefficient λ_U and λ_L depend only on the tail index α and the "correlation" coefficient ρ , precisely $\lambda_U = \lambda_L = \lambda_U(\alpha, \rho)$. Hence the following parametric estimator for the tail-dependence coefficient is suggested:

$$\hat{\lambda}_{U,m}^{(3)} = \hat{\lambda}_{L,m}^{(3)} = \lambda_U^{(3)}(\hat{\alpha}_m, \hat{\rho}_m). \quad (2.10)$$

Several robust estimators $\hat{\rho}_m$ for ρ have been given in the literature (Frahm, Junker, and Schmidt, 2002). Regarding the tail index α , there are well-known estimators obtainable from extreme-value theory. Among these, the Hill estimator represents a natural one for the tail index α :

$$\hat{\alpha}_m = \left(\frac{1}{k} \sum_{j=1}^k \log \|X_{(j,m)}\|_2 - \log \|X_{(k,m)}\|_2 \right)^{-1}, \quad (2.11)$$

where $\|X_{(j,m)}\|_2$ denotes the j -th order statistics of $\|X^{(1)}\|_2, \dots, \|X^{(m)}\|_2$ and

$k = k(m) \rightarrow \infty$ is chosen in an appropriate way. For a discussion on the right choice we refer the reader to Embrechts, Klüppelberg and Mikosch (1997), Drees and Kaufmann (1998), and Groeneboom and de Wolf (2002).

The relationship between the tail index α , the density generator, and the random variable $\|X\|_2$ is established in Schmidt (2002b). Observe that the latter TDC estimator is biased, even though the tail-index estimator is unbiased.

The quantlet `TailCoeffEstimElliptical` estimates the coinciding upper and lower TDC of a bivariate elliptically distributed data-set utilizing the Hill estimator. The inputs of this quantlet are the two-dimensional data-set and the threshold k . The result of the application is assigned to the variable `TDC` and contains the estimated coinciding lower and upper TDC.

```
TDC = TailCoeffEstimElliptical(data,threshold)
      Estimates the upper and lower tail-dependence coefficient based
      on formulas (2.10) and (2.11).
```

2.4 Estimation and empirical results

The figures below reveal that tail dependence is indeed often found in financial data. Provided are two scatter plots of daily negative log-returns of a tuple of financial securities and the corresponding upper TDC estimate $\hat{\lambda}_U^{(1)}$ for various k (for notational convenience we drop the index m). Data-set D_1 contains negative daily stock-log-returns of BMW and Deutsche Bank and data-set D_2 consists of negative daily exchange-rate log-returns of DEM-USD\$ and YEN-USD\$. For modeling reasons we assume that the daily log-returns are iid observations. Both plots show the presence of tail dependence and the order of magnitude of the tail-dependence coefficient. Moreover, the typical variance-bias problem for various threshold values k can be observed. In particular, a small k comes along with a large variance of the TDC estimate, whereas an increasing k results in a strong bias. In the presence of tail dependence, k is chosen such that the TDC estimate $\hat{\lambda}_U^{(1)}$ lies on a plateau between the decreasing variance and the increasing bias. Thus for the data-set D_1 we take k between 80 and 110 which provides a TDC estimate of $\hat{\lambda}_{U,D_1}^{(1)} = 0.28$, whereas for D_2 we estimate $\hat{\lambda}_{U,D_2}^{(1)} = 0.14$.

An application of TDC estimations is given within the Value at Risk (VaR)



Figure 2.4: Scatter plot of BMW versus Dt. Bank negative daily stock log-returns (2325 data points) and the corresponding TDC estimate $\hat{\lambda}_U^{(1)}$ for various k .

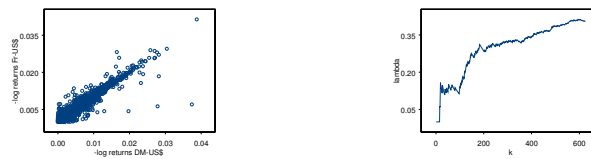


Figure 2.5: Scatter plot of DEM-USD versus YEN-USD negative daily exchange-rate log-returns (3126 data points) and the corresponding TDC estimate $\hat{\lambda}_U^{(1)}$ for various k .

framework of asset portfolios. VaR calculations relate to high quantiles of portfolio-loss distributions and asset return distributions, respectively. In particular, VaR estimations are highly sensitive towards tail behavior and tail dependence of the portfolio's asset-return distributions. Fitting the asset-return random vector towards a multi-dimensional distribution by utilizing a TDC

estimate leads to more accurate VaR estimates. See Schmidt and Stadtmüller (2002) for an extension of the bivariate tail-dependence concept to the multi-dimensional framework. Observe that upper tail dependence for a bivariate random vector $(X_1, X_2)^\top$ is equivalent to

$$\lambda_U = \lim_{\alpha \rightarrow 0} P(X_2 > VaR_{1-\alpha}(X_2) | X_1 > VaR_{1-\alpha}(X_1)) > 0. \quad (2.12)$$

In the last part of this chapter we investigate some properties of the above introduced TDC estimators. 1000 independent copies of $m = 500, 1000, 2000$ iid random vectors of a bivariate standard t -distribution with $\theta = 1.5, 2, 3$ degrees of freedom are generated and the upper TDC's are estimated. The empirical bias and mean-squared-error (MSE) for all three introduced TDC estimation methods are derived and presented in Table 2.4, 2.5, 2.6, 2.7, and 2.8, respectively. For the parametric approach we apply the procedure explained in Section ?? and estimate ρ by a trimmed Pearson-correlation coefficient with trimming portion 0.05% and θ by a Hill-type estimator. Here we choose the optimal threshold value k according to Drees and Kaufmann (1998). Observe that the common "correlation" coefficient ρ does not exist for $\theta \leq 2$. In this case ρ denotes some dependence parameter and a more robust estimation procedure is developed in Frahm, Junker, and Schmidt (2002).

Table 2.4: Bias and MSE for the non-parametric upper TDC estimator $\hat{\lambda}_U^{(1)}$.

Original parameters	$\theta = 1.5$ $\lambda_U = 0.2296$	$\theta = 2$ $\lambda_U = 0.1817$	$\theta = 3$ $\lambda_U = 0.1161$
Estimator	$\hat{\lambda}_U^{(1)}$	$\hat{\lambda}_U^{(1)}$	$\hat{\lambda}_U^{(1)}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0255 (0.00369)	0.0434 (0.00530)	0.0718 (0.00858)
$m = 1000$	0.0151 (0.00223)	0.0287 (0.00306)	0.0518 (0.00466)
$m = 2000$	0.0082 (0.00149)	0.0191 (0.00169)	0.0369 (0.00270)

Table 2.5: Bias and MSE for the non-parametric upper TDC estimator $\hat{\lambda}_U^{(2)}$.

Original parameters	$\theta = 1.5$ $\lambda_U = 0.2296$	$\theta = 2$ $\lambda_U = 0.1817$	$\theta = 3$ $\lambda_U = 0.1161$
Estimator	$\hat{\lambda}_U^{(2)}$	$\hat{\lambda}_U^{(2)}$	$\hat{\lambda}_U^{(2)}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0539 (0.00564)	0.0703 (0.00777)	0.1031 (0.01354)
$m = 1000$	0.0333 (0.00301)	0.0491 (0.00437)	0.0748 (0.00744)
$m = 2000$	0.0224 (0.00173)	0.0329 (0.00228)	0.0569 (0.00436)

Table 2.6: Bias and MSE for the parametric upper TDC estimator $\hat{\lambda}_U^{(3)}$, tail index estimator $\hat{\alpha}$, and $\hat{\rho}$ for $\theta = 1.5$, $\rho = 0$.

Original parameters	$\theta = 1.5$ $\lambda_U = 0.2296$	$\theta = 1.5$ $\alpha = 1.5$	$\theta = 1.5$ $\rho = 0$
Estimator	$\hat{\lambda}_U^{(3)}$	$\hat{\alpha}$	$\hat{\rho}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0016 (0.00093)	0.00224 (0.04492)	0.00002 (0.00379)
$m = 1000$	0.0024 (0.00050)	-0.0147 (0.02354)	-0.0003 (0.00202)
$m = 2000$	0.0024 (0.00024)	-0.0157 (0.01023)	0.00075 (0.00103)

Finally, we provide some plots illustrating the (non-)parametric estimation results for the upper tail-dependence coefficient $\lambda_U^{(i)}$, $i = 1, 2, 3$. Presented are 3×1000 TDC estimations for $m = 500, 1000, 2000$ in a consecutive ordering within three plots for $\theta = 2$. These plots visualize the decreasing empirical variance for increasing m .

Table 2.7: Bias and MSE for the parametric upper TDC estimator $\hat{\lambda}_U^{(3)}$, tail index estimator $\hat{\alpha}$, and $\hat{\rho}$ for $\theta = 2$, $\rho = 0$.

Original parameters	$\theta = 2$ $\lambda_U = 0.1817$	$\theta = 2$ $\alpha = 2$	$\theta = 2$ $\rho = 0$
Estimator	$\hat{\lambda}_U^{(3)}$	$\hat{\alpha}$	$\hat{\rho}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0035 (0.00095)	-0.0198 (0.10417)	-0.00261 (0.00339)
$m = 1000$	0.0058 (0.00057)	-0.0485 (0.05552)	0.0010 (0.00177)
$m = 2000$	0.0054 (0.00029)	-0.0606 (0.02994)	-0.00126 (0.00086)

Table 2.8: Bias and MSE for the parametric upper TDC estimator $\hat{\lambda}_U^{(3)}$, tail index estimator $\hat{\alpha}$, and $\hat{\rho}$ for $\theta = 3$, $\rho = 0$.

Original parameters	$\theta = 3$ $\lambda_U = 0.1161$	$\theta = 3$ $\alpha = 3$	$\theta = 3$ $\rho = 0$
Estimator	$\hat{\lambda}_U^{(3)}$	$\hat{\alpha}$	$\hat{\rho}$
	Bias (MSE)	Bias (MSE)	Bias (MSE)
$m = 500$	0.0162 (0.00115)	-0.2219 (0.29568)	0.0021 (0.00319)
$m = 1000$	0.0154 (0.00076)	-0.2422 (0.20047)	0.0006 (0.00156)
$m = 2000$	0.0124 (0.00046)	-0.2175 (0.12921)	-0.00175 (0.00074)

The empirical studies show that the TDC estimator $\hat{\lambda}_U^{(3)}$ outperforms the other estimators. For $m = 2000$ the bias (MSE) of $\hat{\lambda}_U^{(1)}$ is three (five) times larger than the bias (MSE) of $\hat{\lambda}_U^{(3)}$, whereas the bias (MSE) of $\hat{\lambda}_U^{(2)}$ is two (one and a half) times larger than the bias (MSE) of $\hat{\lambda}_U^{(1)}$. More empirical and statistical results for the estimators $\hat{\lambda}_U^{(1)}$ and $\hat{\lambda}_U^{(2)}$ are given in Schmidt and Stadtmüller (2002). However, remember that the estimator $\hat{\lambda}_U^{(3)}$ was developed for bivariate elliptically-contoured distributions. Thus the estimator $\hat{\lambda}_U^{(1)}$ is recommended for TDC estimations of general or unknown bivariate distributions.

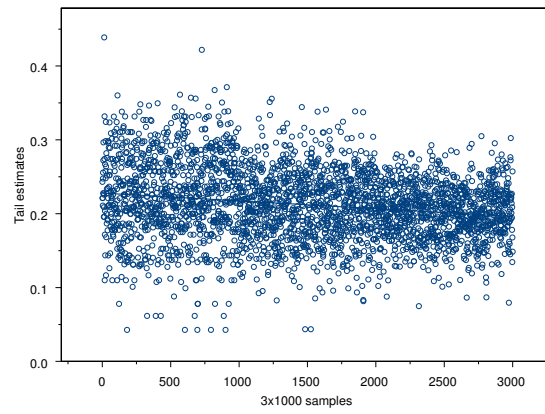


Figure 2.6: Non-parametric upper TDC estimates $\hat{\lambda}_U^{(1)}$ for 3×1000 iid samples of size $m = 500, 1000, 2000$ from a bivariate t -distribution with parameters $\theta = 2$, $\rho = 0$, and $\lambda_U^{(1)} = 0.1817$.

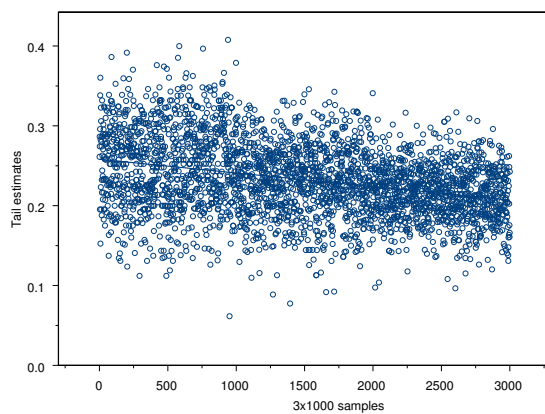


Figure 2.7: Non-parametric upper TDC estimates $\hat{\lambda}_U^{(2)}$ for 3×1000 iid samples of size $m = 500, 1000, 2000$ from a bivariate t-distribution with parameters $\theta = 2$, $\rho = 0$, and $\lambda_U^{(2)} = 0.1817$.

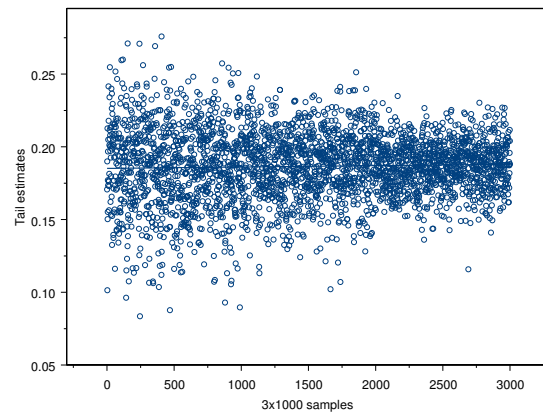


Figure 2.8: Non-parametric upper TDC estimates $\hat{\lambda}_U^{(3)}$ for 3×1000 iid samples of size $m = 500, 1000, 2000$ from a bivariate t -distribution with parameters $\theta = 2$, $\rho = 0$, and $\lambda_U^{(3)} = 0.1817$.

Bibliography

- Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation, Vol. 27 of Encyclopedia of Mathematics and Its Applications*, Cambridge University Press, Cambridge.
- Bingham, N. H. and Kiesel, R. (2002). Semi-parametric modelling in Finance: Theoretical foundation, *Quantitative Finance*, **2**: 241–250.
- Bingham, N. H., Kiesel, R. and Schmidt, R. (2002). Semi-parametric modelling in Finance: Econometric applications, www.mathematik.uni-ulm.de/finmath/.
- Cambanis, S., Huang, S. and Simons, G. (1981). *On the theory of elliptically contoured distributions*, *Journal of Multivariate Analysis* **11**: 368-385.
- Drees, H. and Kaufmann, E. (1998). *Selecting the optimal sample fraction in univariate extreme value estimation*, *Stochastic Processes and their Applications* **75**: 149-172.
- Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, *Applications of Mathematics, Stochastic Modelling and Applied Probability* **33**, Springer Verlag, Berlin.
- Embrechts, P., Lindskog, F. and McNeil, A. (2001). *Modelling Dependence with Copulas and Applications to Risk Management*, ETH preprint.
- Fang, K., Kotz, S. and Ng, K. (1990). *Symmetric Multivariate and Related Distributions*, *Monographs on Statistics and Applied Probability* **36**, Chapman and Hall, London.
- Frahm, G., Junker, M. and Schmidt, R. (2002). *Estimating the Tail Dependence Coefficient*, www.mathematik.uni-ulm.de/finmath/.
- Härdle, W., Kleinow, T. and Stahl, G. (2002). *Applied Quantitative Finance - Theory and Computational Tools e-book*, Springer Verlag, Berlin.

-
- Hauksson, H., Dacorogna, M., Domenig, T., Mueller, U. and Samorodnitsky, G. (2001). *Multivariate Extremes, Aggregation and Risk Estimation*, Quantitative Finance **1**: 79–95.
- Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Monographs on Statistics and Applied Probability **73**, Chapman and Hall, London.
- Ledford, A. and Tawn, J. (1996). *Statistics for Near Independence in Multivariate Extreme Values*, Biometrika **83**: 169–187.
- Nelsen, R. (1999). *An Introduction to Copulas Lecture Notes in Statistics* **139**, Springer Verlag.
- P. Groeneboom, H.L. and de Wolf, P. (2002). *Kernel-type Estimators for the Extreme Value Index*, preprint, Delft University, Netherlands, Retrieved 2002-09-10 from URL ssor.twi.tudelft.nl
- Resnick, S. (2002). *Hidden Regular Variation, Second Order Regular Variation and Asymptotic Dependence*, [www.orie.cornell.edu/~sid/NewForLinking/technical reports.html](http://www.orie.cornell.edu/~sid/NewForLinking/technical%20reports.html).
- Schmidt, R. (2002a). *Credit Risk Modelling and Estimation via Elliptical Copulae*, accepted in "Credit Risk: Measurement, Evaluation and Management", ed. G. Bohl, G. Nakhaeizadeh, and S. Rachev, Springer Verlag.
- Schmidt, R. (2002b). *Tail Dependence for Elliptically Contoured Distributions*, Math. Methods of Operations Research **55**, No. 2: 301–327.
- Schmidt, R. and Stadtmüller, U. (2002). *Non-parametric Estimation of Tail Dependence*, www.mathematik.uni-ulm.de/finmath/.

3 Implied Trinomial Trees

Karel Komorád

The method of Implied Trinomial Trees represents an option pricing technique that tries to fit the market volatility smile. It uses an inductive algorithm constructing a possible evolution process of underlying prices from the current market option data. This chapter recalls the procedure as described in Derman, Kani and Chriss (1996) and shows its general use in [XploRe](#).

3.1 Introduction

In recent decades option-like contracts have been through a giant evolution. Options are financial derivatives speculating on the value of an underlying asset. The boom in research on options and the use of them started after Black and Scholes published the formula for options prices. Options are maybe more common than one could think – many issues of corporate securities (e.g. bonds or stocks) also have option features. New financing techniques, for instance, contingent value rights (CVR's), are straightforward applications of options. Liquidly traded options are important financial instruments used for **hedging** – they can be included into the portfolio to reduce risk. So option pricing has become one of the basic techniques in finance.

Unfortunately, option prices computed by the Black-Scholes formula and the market options prices show a discrepancy. This deviation may be due to the assumption of constant volatility in the Black-Scholes model which is hardly ever fulfilled in the market. Therefore, new approaches were examined to find options prices consistent with the market. Binomial trees, discrete versions of the Black-Scholes model, are intuitive and probably the most commonly used basic procedure for option pricing. Derman and Kani (1994) presented an extension of binomial trees called implied binomial trees that fit the market data. Implied trinomial trees (ITT) are analogical extension of trinomial trees.

Trinomial trees converge in the continuous limit to the same result as binomial trees, but the use of ITT's give us a freedom in the choice of the possible path of the underlying price, the so called **state space**. This freedom could be in some circumstances desirable. Figure ??(a) shows a possible state space with three periods over six years. An ITT with one-year time steps over ten years is depicted in Figure ??(b).

3.2 Basic Option Pricing Overview

Options are contingent claims on the value of an underlying asset, usually a stock or a traded market index. The simplest type of options are **European options**. Their importance is mainly theoretical. An European **call** (or **put**) gives the owner the right to buy (or sell) an underlying asset at a fixed price (the so called **strike** price) at a given date.

The famous Black-Scholes model of option pricing can be viewed as a continuous generalization of duplication model. For an explanation of the duplication model see Ross, Westerfield and Jaffe (2002). Black and Scholes assume that the underlying asset follows a geometric Brownian motion with a constant volatility and is described by the stochastic differential equation

$$\frac{dS_t}{S_t} = \mu dt + \sigma dZ_t, \quad (3.1)$$

where S denotes the underlying price, μ is the expected return and Z stands for the standard Wiener process. As a consequence, the underlying distribution is lognormal. This is the crucial point – volatility is the only parameter of the Black-Scholes formula which is not explicitly observable from the market data. When plugging the market option prices into the formula and solving for volatility (the so called **implied volatility**), empirical data shows a variation with both exercise (the **skew** structure) and expiration time (the **term** structure), both together are commonly called the **volatility smile**. Figure 3.1 illustrates this behavior. Panel 3.1(a) shows the decrease of implied volatility σ_{imp} with the strike level of the put options on the DAX index with a fixed expiration of 52 days, as observed on January 29, 1999. Panel 3.1(b) shows the increase of σ_{imp} with the time to expiration of put options with a fixed strike level 6000 DM.

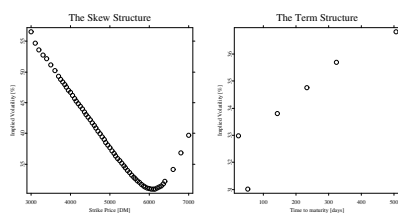


Figure 3.1: Implied Volatilities of DAX Options on Jan 29, 1999.

Implied volatility of an option is the market's estimate of the average future underlying volatility during the life of that option, **local volatility** is the market's estimate of underlying volatility at a particular future time and market level. Let us recall that as a consequence of complete markets, the Black-Scholes model is risk-neutral.

The most natural simplification of the Black-Scholes model are binomial trees. The binomial tree corresponding to the risk neutral underlying evaluation process is the same for all options on this asset, no matter what the strike price or time to expiration is. The tree cannot know which option we are valuating on it. There are several alternative ways to construct a binomial tree. Here we will recall the familiar Cox-Ross-Rubinstein (CRR) binomial tree. It has a constant logarithmic spacing between nodes on the same level. This spacing shows the future price volatility. Figure 3.2 shows a standard CRR tree. Starting at a general node S , the price of the underlying asset can either increase to the higher price S_u with the probability p or decrease to the lower price S_d with the probability $1 - p$:

$$S_u = Se^{\sigma\sqrt{\Delta t}} \quad (3.2)$$

$$S_d = Se^{-\sigma\sqrt{\Delta t}} \quad (3.3)$$

$$p = \frac{F - S_d}{S_u - S_d} \quad (3.4)$$

where F denotes the forward price in the node S and σ is the constant volatility. For other types of binomial trees see Derman, Kani and Chriss (1996) .

New models try to overcome the problem of varying volatility and to price options consistent with the market prices – to account for the volatility smile. However many of them extend the original Black-Scholes theory. Let us mention, for instance, incorporating a stochastic volatility factor or discontinuous jumps in the underlying price. But these methods bring other problems. In the next section we will pay attention to another approach of Black-Scholes extension developed in Derman and Kani (1994) – to the implied trees.

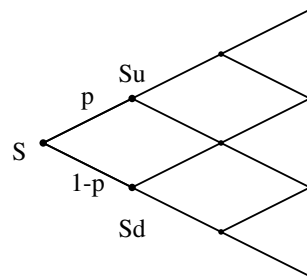


Figure 3.2: CRR Binomial Tree.

3.3 Trees and Implied Models

While the Black-Scholes attitude assumes that the underlying asset follows geometric Brownian motion from equation (??) with a constant volatility, implied theories assume that the stock or index price follows a process, whose volatility $\sigma(S, t)$ varies with the spot price and time. This process can be expressed by the following stochastic differential equation:

$$\frac{dS_t}{S_t} = \mu dt + \sigma(S, t) dZ_t. \quad (3.5)$$

This approach ensures that the valuation of the option remains preference-free (all uncertainty is in the spot price and thus we can hedge options using the underlying). Derman and Kani (1994) shows that it is possible to determine $\sigma(S, t)$ directly from the market prices of liquidly traded options.

The most natural discrete representation of a non-lognormal evolution process of the underlying prices is an implied binomial tree (IBT). It is similarly like the

CRR binomial tree is a discrete version of the Black-Scholes constant-volatility process. For an exhaustive explanation on IBT's see Derman and Kani (1994). Generally we can use any (higher) multinomial tree as a discrete development from the Black-Scholes model. However, all of them converge, as the time step tends towards zero, to the same continuous result – the constant-volatility process. IBT's are, from the set of all implied multinomial trees, minimal, i.e. they have only one degree of freedom – the arbitrary choice of the central node at each level of the tree. In this sense one may feel that they are sufficient and no more complicated trees are necessary. Despite the limiting similarity of multinomial trees, one form of them could be more convenient. Sometimes one may wish more flexibility in the discrete world – to get transition probabilities and probability distributions which vary as smoothly as possible across the tree. This is important especially when the market option prices are not very precise (e.g. because of inefficiency, market frictions, etc.). It is useful to recall the concept of Arrow-Debreu prices before the following derivation in Section 3.4. **Arrow-Debreu** price $\lambda_{n,i}$ at node (n, i) is computed as the sum over all paths starting in the root of the tree leading to node (n, i) of the product of the risklessly discounted transition probabilities. Arrow-Debreu price of the root is equal to one. Arrow-Debreu prices at the final level of a (multinomial) tree are a discrete approximation of the implied distribution. Notice that they are discounted and thus the risk-neutral probability corresponding to each node (at the final level) should be calculated as the product of the Arrow-Debreu price and the factor e^{rT} .

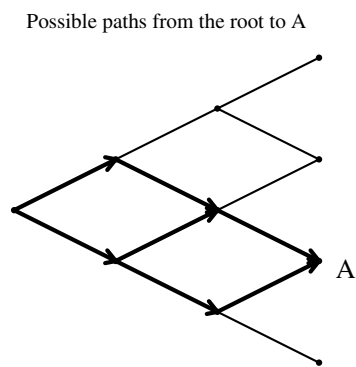


Figure 3.3: Computing Arrow-Debreu Price in a Binomial Tree.

3.4 ITT's and Their Construction

3.4.1 Basic insight

A trinomial tree with N levels is a set of nodes $s_{n,i}$, $n = 1, \dots, N$, $i = 1, \dots, 2n-1$. The underlying price can move from each node $s_{n,i}$ to one of three nodes: with probability p_i to the upper node, value $s_{n+1,i}$, with probability q_i to the lower node, value $s_{n+1,i+2}$ and with probability $1 - p_i - q_i$ to the middle node, value $s_{n+1,i+1}$. Let us denote the nodes in the new level with capital letters: $S_i (= s_{n+1,i})$, $S_{i+1} (= s_{n+1,i+1})$ and $S_{i+2} (= s_{n+1,i+2})$ respectively.

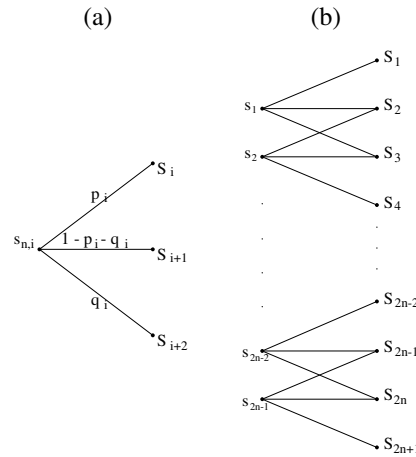


Figure 3.4: General Nodes in a Trinomial Tree.

Figure 3.4(a) shows a single node of a trinomial tree. Starting in the node $s_{n,i}$ at time t_n there are five unknown parameters: the two transition probabilities $p_{n,i}$, $q_{n,i}$ and the three new node prices S_i , S_{i+1} , and S_{i+2} . Let F_i denote the known forward price of the spot $s_{n,i}$ and λ_i the known Arrow-Debreu price at node (n, i) . In particular, Arrow-Debreu prices for a trinomial tree can be

obtained by the following iterative formulas:

$$\lambda_{1,1} = 1 \quad (3.6)$$

$$\lambda_{n+1,1} = e^{-r\Delta t} \lambda_{n,1} p_{n,1} \quad (3.7)$$

$$\lambda_{n+1,2} = e^{-r\Delta t} (\lambda_{n,1}(1 - p_{n,1} - q_{n,1}) + \lambda_{n,2} p_{n,2}) \quad (3.8)$$

$$\lambda_{n+1,i+1} = e^{-r\Delta t} (\lambda_{n,i-1} q_{i-1} + \lambda_{n,i}(1 - p_{n,i} - q_{n,i}) + \lambda_{n,i+1} p_{n,i+1}) \quad (3.9)$$

$$\lambda_{n+1,2n} = e^{-r\Delta t} (\lambda_{n,2n-1}(1 - p_{n,2n-1} - q_{n,2n-1}) + \lambda_{n,2n-2} q_{n,2n-2}) \quad (3.10)$$

$$\lambda_{n+1,2n+1} = e^{-r\Delta t} \lambda_{n,2n-1} q_{n,2n-1} \quad (3.11)$$

Figure 3.4(b) shows two levels of a general trinomial tree ¹. The evolution process of underlying prices can be seen discrete through an implied binomial tree. Implied trinomial tree (ITT) is another discrete representation of the same process. What we desire from an ITT, and in general from any implied tree, is:

1. the tree reproduces correctly the volatility smile,
2. the tree is risk-neutral,
3. all the transition probabilities are from the interval $(0, 1)$.

To fulfill the risk-neutrality condition, the expected value of the underlying price $s_{n,i}$ in the next period t_{n+1} must be its known forward price:

$$E s_{n,i} = p_i S_i + (1 - p_i - q_i) S_{i+1} + q_i S_{i+2} = F_i = e^{(r-\delta)\Delta t} s_{n,i}, \quad (3.12)$$

where r denotes the continuous interest rate, δ the dividend yield and Δt is the time step from t_n to t_{n+1} . Derman, Kani and Chriss (1996) say that the node prices and transition probabilities also satisfy:

$$p_i (S_i - F_i)^2 + (1 - p_i - q_i) (S_{i+1} - F_i)^2 + q_i (S_{i+2} - F_i)^2 = F_i^2 \sigma_i^2 \Delta t + o(\Delta t), \quad (3.13)$$

where σ_i is the stock or index price volatility during this time period and $o(\Delta t)$ denotes a term of higher order than Δt . Therefore we have two constraints in equations (3.12) and (3.13) for five unknown parameters. As a consequence, there is no unique implied trinomial tree, we can thus choose one of the large set of equivalent ITT's. Equivalent in the sense, that as the time spacing Δt tends towards zero, all of these trees tend to give the same results. A common

¹Here we slightly modified the notation of Derman, Kani and Chriss (1996) to be consistent with the program code.

method is to choose freely the underlying prices and then to use these two equations and solve them for the transition probabilities p_i, q_i . Afterward we must make sure that these probabilities do not violate the condition 3 above. Since in general all multinomial trees converge to the same theoretical result (namely to the continuous process of evolution in the underlying asset), the reason why we use ITT's instead of IBT's is that we gain these three degrees of freedom. This freedom may allow us to fit the smile better, especially when inconsistent or arbitrage-violating market options prices make a consistent tree impossible. Even though the constructed tree is consistent, other difficulties can arise when its local volatility and probability distributions are jagged and "implausible".

3.4.2 State space

There are several methods we can use to construct an initial state space. As already mentioned, all of them converge to the same theory (constant-volatility Black-Scholes model), in the continuous limit. In other words all these models are equivalent discretizations of the constant volatility diffusion process. We chose a constant-volatility Cox-Ross-Rubinstein (CRR) binomial tree. Then we combined two steps of this tree into a single step of our new trinomial tree. This is illustrated in Figure 3.5. Using equations (3.2) and (3.3) we can derive the following formulas for the nodes in the new trinomial tree:

$$S_i = s_{n+1,i} = s_{n,i} e^{\sigma\sqrt{2\Delta t}} \quad (3.14)$$

$$S_i + 1 = s_{n+1,i+1} = s_{n,i} \quad (3.15)$$

$$S_i + 2 = s_{n+1,i+2} = s_{n,i} e^{-\sigma\sqrt{2\Delta t}} \quad (3.16)$$

Now, summing up the probabilities of reaching one of the three nodes, using formula (3.4), we get the relationship for the up and down transition probabilities in the trinomial tree (the middle transition probability is equal to $1 - p_i - q_i$):

$$p_i = \left(\frac{e^{r\Delta t/2} - e^{-\sigma\sqrt{\Delta t/2}}}{e^{\sigma\sqrt{\Delta t/2}} - e^{-\sigma\sqrt{\Delta t/2}}} \right)^2$$

$$q_i = \left(\frac{e^{\sigma\sqrt{\Delta t/2}} - e^{r\Delta t/2}}{e^{\sigma\sqrt{\Delta t/2}} - e^{-\sigma\sqrt{\Delta t/2}}} \right)^2$$

Derman, Kani and Chriss (1996) describes two other methods for building a constant-volatility trinomial tree.

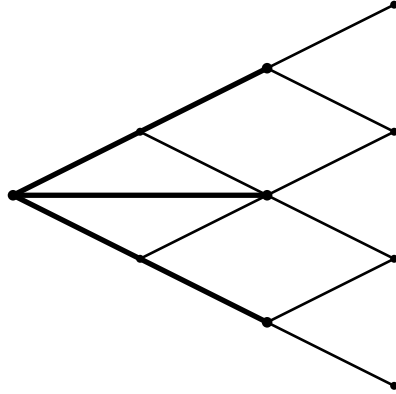


Figure 3.5: Building a Constant-Volatility Tree by Combining Two Steps of a CRR Binomial Tree.

When the implied volatility varies only slowly with strike and expiration, the regular state space with uniform mesh size is adequate for constructing ITT models. But if the volatility varies significantly with strike or time to maturity, we should choose a state space reflecting this properties. Construction of a trinomial space with proper skew and term structure proceeds in four steps.

- First, we build a regular trinomial lattice with constant time spacing Δt and constant level spacing, assuming all interest rates and dividends equal to zero. Such a lattice can be constructed in the way described above.
- Secondly, we modify Δt at different times. Let us denote the known equally spaced time points $t_0 = 0, t_1, \dots, t_n = T$, then we can find the unknown scaled times $\tilde{t}_0 = 0, \tilde{t}_1, \dots, \tilde{t}_n = T$ by solving the following set

of non-linear equations:

$$\tilde{t}_k \sum_{i=1}^{n-1} \frac{1}{\sigma^2(\tilde{t}_i)} + \tilde{t}_k \frac{1}{\sigma^2(T)} = T \sum_{i=1}^k \frac{1}{\sigma^2(\tilde{t})} \quad ; k = 1, \dots, n-1. \quad (3.17)$$

- Subsequently, we change ΔS at different levels. Denoting S_1, \dots, S_{2n+1} the original known underlying prices and $\tilde{S}_1, \dots, \tilde{S}_{2n+1}$ the new scaled unknown underlying prices respectively, we solve the following equation:

$$\frac{\tilde{S}_k}{\tilde{S}_{k-1}} = \exp \left\{ \frac{c}{\sigma(S_k)} \ln \frac{S_k}{S_{k-1}} \right\} \quad k = 2, \dots, 2n+1 \quad (3.18)$$

for some constant c (recommended to be some typical value of the local volatility). There are $2n$ equations for $2n+1$ unknown parameters. Here we always suppose that the new central node is the same as the original central node: $\tilde{S}_{n+1} = S_{n+1}$. For a more elaborate explanation of the theory behind equations (3.17) and (3.18), set in the continuous frame, see Derman, Kani and Chriss (1996).

- Lastly, you can grow all the node prices by a sufficiently large growth factor which removes forward prices violations (explained in Subsection 3.4.4). Multiplying all zero-rate node prices at time t_n by $e^{(r-\delta)(t_n-t_{n-1})}$ should be always sufficient.

3.4.3 Transition probabilities

Once the state space of the ITT is fixed, we can compute the transition probabilities for all tree nodes (n, i) at each tree level n . Let $C(K, t_{n+1})$ and $P(K, t_{n+1})$ denote today's price for a standard European call and put option, respectively, struck at K and expiring at t_{n+1} . These values can be obtained by interpolating the smile surface at various strike and time points. Then the trinomial tree value of an option is the sum over all nodes $(n+1, j)$ of the discounted probability of reaching that node multiplied by the pay-off function, i.e. $\max(S_j - K, 0)$ for the call and $\max(K - S_j, 0)$ for the put option respectively, where S_j denotes the spot price at the node $(n+1, j)$:

$$C(K, t_{n+1}) = e^{-r\Delta t} \sum_j \{p_j \lambda_j + (1 - p_{j-1} - q_{j-1}) \lambda_{j-1} + q_{j-2} \lambda_{j-2}\} (S_j - K)^+ \quad (3.19)$$

$$P(K, t_{n+1}) = e^{-r\Delta t} \sum_j \{p_j \lambda_j + (1 - p_{j-1} - q_{j-1}) \lambda_{j-1} + q_{j-2} \lambda_{j-2}\} (K - S_j)^+ \quad (3.20)$$

If we set the strike price K to the value S_{i+1} , the stock price at the node $(n+1, i+1)$, rearrange the terms in the sum and use equation (3.12) we can compute from equation (3.19) the transition probabilities p_i, q_i for all the nodes above the central node:

$$p_i = \frac{e^{r\Delta t} C(S_{i+1}, t_{n+1}) - \sum_{j=1}^{i-1} \lambda_j (F_j - S_{i+1})}{\lambda_i (S_i - S_{i+1})} \quad (3.21)$$

$$q_i = \frac{F_i - p_i (S_i - S_{i+1}) - S_{i+1}}{S_{i+2} - S_{i+1}} \quad (3.22)$$

Similarly, we compute the transition probabilities from equation (3.20) for all the nodes below (and including) the center node $(n+1, n)$ at time t_n :

$$q_i = \frac{e^{r\Delta t} P(S_{i+1}, t_{n+1}) - \sum_{j=i+1}^{2n-1} \lambda_j (S_{i+1} - F_j)}{\lambda_i (S_{i+1} - S_{i+2})} \quad (3.23)$$

$$p_i = \frac{F_i - q_i (S_{i+2} - S_{i+1}) - S_{i+1}}{S_i - S_{i+1}} \quad (3.24)$$

Derivation of these equations is described in detail in Komorád (2002). Finally the implied local volatilities are approximated from equation (3.13):

$$\sigma_i^2 \doteq \frac{p_i (S_i - F_i)^2 + (1 - p_i - q_i) (S_{i+1} - F_i)^2 + q_i (S_{i+2} - F_i)^2}{F_i^2 \Delta t}. \quad (3.25)$$

3.4.4 Possible pitfalls

Equations (3.21) - (3.24) can unfortunately result in transition probabilities which are negative or greater than 1. This is inconsistent with rational option prices and allows arbitrage. We must face two forms of this problem. First, we must watch that no forward price $F_{n,i}$ of a node (n, i) falls outside the range of daughter nodes at the level $n+1$: $F_{n,i} \in (s_{n+1, i+2}, s_{n+1, i})$. Figure 3.6 shows this type of problem. This is not very difficult to overcome, since the state

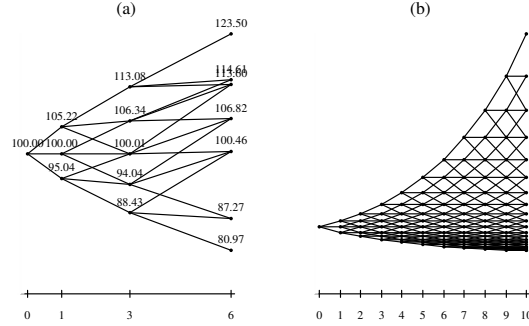


Figure 3.6: Two Kinds of the Forward Price Violation.

space is our free choice – we can overwrite the nodes causing this problem. The second problem refers to extreme values of options prices, which would imply an extreme value of local volatility. This may not be possible to obtain with probabilities between $(0, 1)$. In such case we have to overwrite the option price which produces the unacceptable probabilities. In practice it means that we overwrite even these probabilities with numbers from the interval $(0, 1)$. This is always possible if our state space does not violate the forward condition. From the great number of possible ways to overwrite wrong probabilities with numbers between 0 and 1, we chose to set:

$$p_i = \frac{1}{2} \left(\frac{F_i - S_{i+1}}{S_i - S_{i+1}} + \frac{F_i - S_{i+2}}{S_i - S_{i+2}} \right), \quad q_i = \frac{1}{2} \left(\frac{S_i - F_i}{S_i - S_{i+2}} \right) \quad (3.26)$$

if $F_i \in (S_{i+1}, S_i)$ and

$$p_i = \frac{1}{2} \left(\frac{F_i - S_{i+2}}{S_i - S_{i+2}} \right), \quad q_i = \frac{1}{2} \left(\frac{S_{i+1} - F_i}{S_{i+1} - S_{i+2}} + \frac{S_i - F_i}{S_i - S_{i+2}} \right) \quad (3.27)$$

if $F_i \in (S_{i+2}, S_{i+1})$. In either case the middle transition probability is equal to $1 - p_i - q_i$.

3.4.5 Illustrative examples

To illustrate the construction of an implied trinomial tree, let us now consider that the volatility varies only slowly with the strike price and time to expiration.

Assume that the current index level is 100 EUR, the annual riskless interest rate $r = 12\%$, dividend yield is $\delta = 4\%$. The annual Black-Scholes implied volatility is assumed to be $\sigma = 11\%$, moreover let us assume that the Black-Scholes implied volatility increases (decreases) linearly by 10 basis points ($=0.1\%$) with every 10 unit drop (rise) in the strike price, i.e. $\sigma_{impl} = 0.11 - \Delta Strike * 0.001$. To keep the example easy, we consider one-year steps over three years. First, we construct the state space – a constant-volatility trinomial tree as described in Section 3.4.2. The first node at time $t_0 = 0$, labeled A in Figure 3.7, has

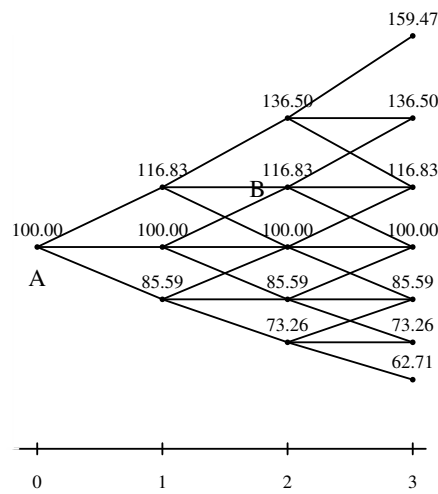


Figure 3.7: State Space of a Trinomial Tree with Constant Volatility of 11%.

the value of $s_A = 100$ – today's spot price. The next three nodes, at time t_1 , are computed from equation (3.14): $S_1 = 116.83$, $S_2 = 100.00$ and $S_3 = 85.59$, respectively. In order to determine the transition probabilities, we need to know the put price struck at $S_2 = 100$ and expiring one year from now $P(S_2, t_1)$. From the smile, the implied volatility of this option is 11%. We calculate its price using a constant-volatility trinomial tree with the same state space, and find it to be 0.987 EUR. The forward price corresponding to node A is $F_A = S e^{(r^* - \delta^*)\Delta t} = 107.69$. Here r^* and δ^* denote the continuous interest

rate $r^* = \log(1 + r)$ and $\delta^* = \log(1 + \delta)$, respectively. The down transition probability is then computed from equation (3.23)

$$q_A = \frac{e^{\log(1+0.12) \times 1} 0.987 - \Sigma}{1 \times (100.00 - 85.59)} = 0.077.$$

The summation term Σ in the numerator is zero in this case, because there are no nodes with price lower than S_3 at time t_1 . The up transition probability p_A is computed from equation (3.24)

$$p_A = \frac{107.69 + 0.077 \times (100.00 - 85.59) - 100}{116.83 - 100.00} = 0.523.$$

The middle transition probability is then equal to $1 - p_A - q_A = 0.6$. As one can see from equations (3.6) – (3.11) the Arrow-Debreu prices turn out to be just discounted transition probabilities: $\lambda_{1,1} = 0.467$, $\lambda_{1,2} = 0.358$ and $\lambda_{1,3} = 0.069$. Equation (3.25) then gives the value of the implied local volatility at node A: $\sigma_A = 9.5\%$. Let us show the computation of one node more. Have a look at the node B in year 2 of Figure 3.7. The index level at this node is $s_B = 116.83$ and its forward price one year later is $F_B = 125.82$. From this node, the index can move to one of three future nodes at time $t_3 = 3$, with prices $s_{3,2} = 136.50$, $s_{3,3} = 116.83$ and $s_{3,4} = 100.00$, respectively. The value of a call option struck at 116.83 and expiring at year 3 is $C(s_{3,3}, t_3) = 8.87$, corresponding to the implied volatility of 10.83% interpolated from the smile. The Arrow-Debreu price $\lambda_{2,2}$ is computed from equation (3.8):

$$\lambda_{2,2} = e^{-\log(1+0.12) \times 1} (0.467 \times (1 - 0.517 - 0.070) + 0.358 \times 0.523) = 0.339.$$

The numerical values here are the already known values from the last level t_1 . The complete trees of Arrow-Debreu prices, transition probabilities and local volatilities for our example are shown in Figures 3.8 – 3.10. Now, using equations (3.21) and (3.22) we can find the transition probabilities:

$$\begin{aligned} p_{2,2} &= \frac{e^{\log(1+0.12) \times 1} \times 8.87 - \Sigma}{0.339 \times (136.50 - 116.83)} = 0.515 \\ q_{2,2} &= \frac{125.82 - 0.515 \times (136.50 - 116.83) - 116.83}{100 - 116.83} = 0.068, \end{aligned}$$

where Σ contributes with the only term $0.215 \times (147 - 116.83)$ (there is one single node above S_B whose forward price is equal to 147). Last from equation (3.25), we find the implied local volatility at this node $\sigma_B = 9.3\%$.

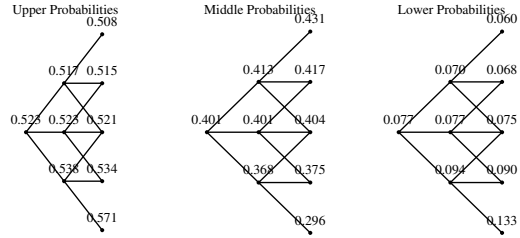


Figure 3.8: Transition Probabilities when $\sigma_{impl} = 0.11 - \Delta Strike * 0.001$.

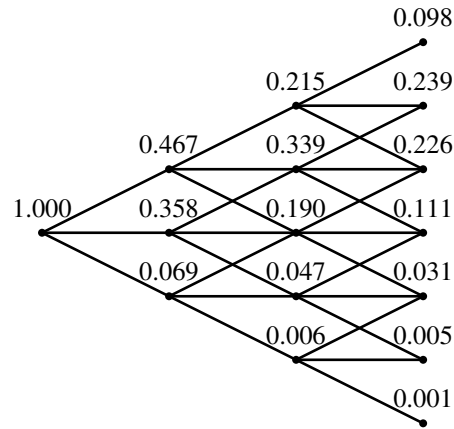


Figure 3.9: Tree of Arrow-Debreu prices when $\sigma_{impl} = 0.11 - \Delta Strike * 0.001$.

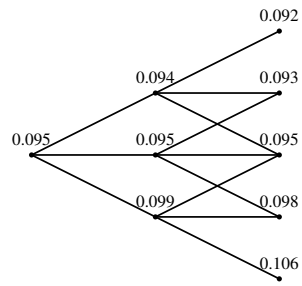


Figure 3.10: Implied Local Volatilities when $\sigma_{impl} = 0.11 - \Delta Strike * 0.001$.

As already mentioned in Section 3.4.4, the transition probabilities may fall out of the interval $(0, 1)$. If we slightly modify our previous example, we can show such a case. Let us now assume that the Black-Scholes volatility from our example increases (decreases) linearly 0.5 percentage points with every 10 unit drop (rise) in the strike price, i.e. $\sigma_{impl} = 0.11 - \Delta Strike * 0.05$ (that is, the volatility is five times as steep as before – illustrated in Figure 3.11). Using the same state space, we find unadmissible transition probabilities at nodes C and D of Figures 3.12 – 3.14. To overwrite them with plausible values, we used the set up described in equations (3.26) and (3.27).

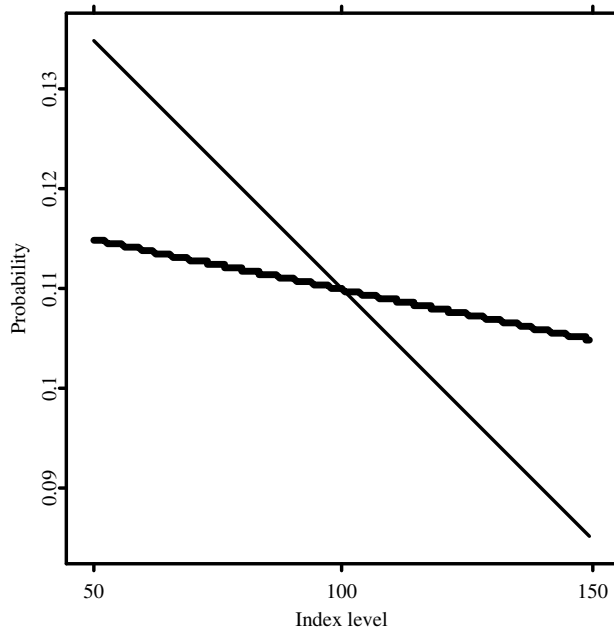


Figure 3.11: Skew structure from the examples. Thick line for $\sigma_{impl} = 0.11 - \Delta Strike * 0.001$ and thin line when $\sigma_{impl} = 0.11 - \Delta Strike * 0.005$.

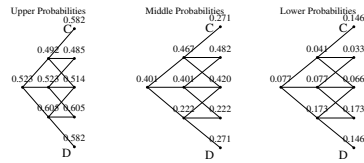


Figure 3.12: Transition Probabilities when $\sigma_{impl} = 0.11 - \Delta Strike * 0.005$.

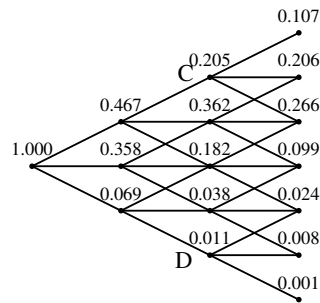


Figure 3.13: Tree of Arrow-Debreu prices when $\sigma_{impl} = 0.11 - \Delta Strike * 0.005$.

3.5 Computing Implied Trinomial Trees

3.5.1 Basic skills

The simplest way to invoke implied trinomial trees in [XploRe](#) is to use the interactive menus, which guide you through the computation procedure:

```
{Ttree, P, Q, AD, LocVol, Onodes, Oprobs, Time} = ITT (volafunc)
```

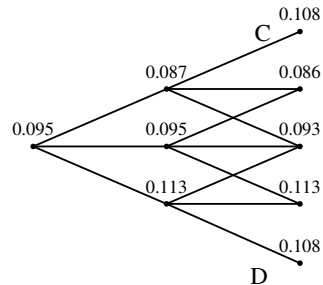


Figure 3.14: Implied Local Volatilities when $\sigma_{impl} = 0.11 - \Delta Strike * 0.005$.

There is only one parameter that has to be inserted – the name of the volatility function(s) `volafunc`. The output parameters will be discussed later on. First, we consider only slow volatility variation. Then `volafunc` is one string containing the name of the volatility function defined by the user. This function has the following form:

```
sigma = volafunc (S, K, tau)
```

where `S` denotes the spot price, `K` the exercise price and `tau` is the time to maturity on each level. For our first example we will assume that the implied volatility is 15% for all expirations and that it increases (decreases) 0.5% with every 10 points drop (rise) in the strike price:

```
library("finance")           ; load the library
proc(sigma)=volafunc(S, K, tau) ; define the volatility function
  sigma=0.015 + (S-K)/10 * 0.05
endp
ITT("volafunc")             ; ITT's with interactive menus
                             ■ STFitt01.xpl
```

XploRe asks you to insert the strike price, the annual interest rate and the divided yield. Subsequently, you must insert all the time points, when the steps of the ITT occurred. At this stage let us recall that the tree is supposed to start at time $t_0 = 0$. To end the insertion of the time steps, type any non-positive value. For instance, in our example from subsection 3.4.5, we typed consecutively: 0, 1, 2, 3 and 0. Once the ITT is computed, you can display the results – a menu asks you to choose between the state space, the tree of the transition probabilities, the tree of Arrow-Debreu prices, the tree of local volatilities and the state price density. If you choose the state space of the tree you can specify whether you want to describe the nodes with underlying prices or the arrows with transition probabilities respectively. While in this plot the scale of the y-axis corresponds to the underlying price, the tree in any other graph is only auxiliary and the vertical spacing does not mean anything.

Next, let us consider significant term and skew structure. In such cases the input parameter `volafunc` is a string vector with three rows containing names of functions with the following syntax (the names are arbitrary):

```
sigma = term (t)
sigmader = termder (t)
sigma = skew (S)
```

The first row contains the name of the volatility function for the term structure defined by the user. The second row is the name of the quantlet returning Jacobian of this term structure function. Both of these two quantlets have only one parameter time t and it is necessary that they can use vectors. The third row is the name of the function for the skew structure. It has one input parameter S which is the spot price. If there is no term structure, but significant skew structure, insert in the first row only an empty string: `volafunc[1]=""`. The following quantlet computes the state space of an ITT with significant term and skew structure.

■ STFitt02.xpl

The state space is plotted in Figure 3.15. Now we hid the part which manages about the plotting routines, this will be explained later.

The general form of ITT without the interactivity is

```
{Ttree, P, Q, AD, LocVol, Onodes, Oprobs, Time} = ITT(S,
  r, div,
  time, volafunc{, skewconst})
```

where S is the spot price of the underlying at time $t_0 = 0$, r and div are compounded riskless interest rate and dividend yield (both of them $\in (0, 1)$). $time$ is vector of time points corresponding to the tree. They are to be given in years and they are supposed to increase. The parameter $volafunc$ has been already described above and $skewconst$ is an optional parameter which will be discussed in Section 3.5.2. The output $Ttree$ is a matrix containing the Implied Trinomial Tree. $Ttree[1,1]$ is the root and all elements which do not belong to the tree are NaN's. P and Q are matrices with the up and down transition probabilities respectively, AD is a matrix containing the Arrow-Debreu prices for the computed tree and $LocVol$ is a matrix with implied local volatilities. If there were some violations of the forward condition (described in Subsection 3.4.4) then $Onodes$ is a matrix of NaN's with the same dimensions like $Ttree$, but on the places of the overwritten nodes is their value. If there were no overwritten nodes $Onodes$ is only scalar ($=NaN$). $Oprobs$ is a matrix with two columns containing x- and y-coordinates of the nodes, where a probability violated the second condition from Section 3.4.4. If there were none, $Oprobs$ is equal to NaN. But for each probability causing an overwrite, four rows are appended to the first auxiliary row $0\sim 0$. First row of this quartet is the parental node where overwrite occurred, the three other rows are corresponding daughter nodes. $Time$ is the same as the input parameter $time$, if there is no significant term structure. In other case it contains the solution of the equation system 3.17. So if we change the parameter $time$ from our last example: $time=0|1|2$, then the variable `impltree` will contain these following components:

Table 3.1: Contents of `impltree.Ttree`

100	110.5	119.94
+NAN	103.97	112.18
+NAN	97.461	105.56
+NAN	+NAN	98.944
+NAN	+NAN	91.247

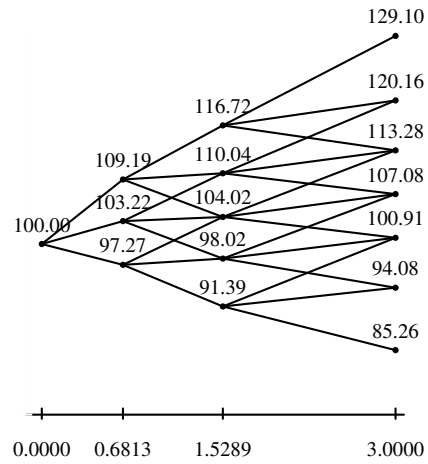


Figure 3.15: State Space with Significant Term and Volatility Structure.

Table 3.2: Contents of impltree.P

0.45353	0.67267
+NAN	0.72456
+NAN	0.70371

Table 3.3: Contents of impltree.Q

0.00053582	0.11468
+NAN	0.091887
+NAN	0.093641

Table 3.4: Contents of `impltree.AD`

1	0.4123	0.24826
+NAN	0.4963	0.40038
+NAN	0.00048711	0.12417
+NAN	+NAN	0.04091
+NAN	+NAN	4.0831e-05

Table 3.5: Contents of `impltree.LocVol`

0.049924	0.040269
+NAN	0.036105
+NAN	0.04117

Table 3.6: Contents of `impltree.Onodes`

+NAN

The simplest way to display results is to use quantlet `plotITT`. It specifies explicitly which components of the computed ITT should be plotted.

```
plotITT (itree, what{, r{, text{, prtext}}})
```

The input parameter `itree` is just the output parameter of the quantlet `ITT`, i.e. a list containing `Ttree`, `P`, `Q`, `AD`, `LocVol`, `Onodes`, `Oprobs` and `Time`. `what` is a vector determining, which plots will be displayed. It has up to 5 rows and each non-zero tells `XploRe` that the corresponding plot should be shown. The possible plots are in sequence: 1) state space of the given ITT, 2) trees with transition probabilities, 3) the tree of local volatilities, 4) the tree of Arrow-Debreu prices and 5) the state price density. `r` is the interest rate ($\in (0, 1)$), and

Table 3.7: Contents of impltree.Oprobs

0	0
0.83772	110.5
2	119.94
2	112.18
2	105.56
0.83772	103.97
2	112.18
2	105.56
2	98.944
0.83772	97.461
2	105.56
2	98.944
2	91.247

Table 3.8: Contents of impltree.Time

0
0.83772
2

it must be included whenever you want to show the state price density. `text` is an optional parameter which can be used only when the state space of the ITT should be plotted. It says whether the tree nodes should be described (`text` is a nonzero) or not (`text = 0`). And similarly `prtext` determines whether the arrows connecting the state space should be described with corresponding transition probabilities. In default, the state space denotes any node which violated the forward condition with a red rhombus and overwritten probabilities are shown with magenta lines. For instance, the following commands produce plot of the state price density of the implied trinomial tree from Figure ??(b). The state price density is depicted in Figure 3.16.

■ STFitt04.xpl

3.5.2 Advanced features

We have mentioned a constant c in Section 3.4.2. Derman, Kani and Chriss (1996) suggest to set this constant, used for solving the non-linear equation system (3.18), to represent the at-the-money or some other typical value of local volatility. In default it is set to 0.1 in the quantlet ITT, but the parameter `skewconst` allows to set it for each time step on a different value.

In order to allow the user to compose the plot of any trinomial tree (it does not matter if it is a tree of state space, transition probabilities or something else) with other graphical objects into one graph, there are quantlets returning the tree as a graphical object. Moreover these quantlets allow the user to enjoy other features by plotting the trees. Let us start with the quantlet `grITTspd`:

```
dat = grITTspd (ttree,ad,r,time)
```

Here `ttree`, `ad` and `time` are the output parameters `Ttree`, `AD` and `Time` respectively, which were returned by the quantlet `ITT`; `r` is then the corresponding compound interest rate. The output parameter `dat` is the estimated state price density as a graphical object. The density curve will be estimated using quantlet `regest` with quartic kernel and bandwidth of 20% of the range of the last level in `ttree`.

Next, we introduce a quantlet used in general for any component of a trinomial tree:

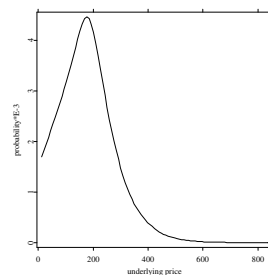


Figure 3.16: Estimation of the State Price Density from an ITT.

```
nodes = grITTCrr (onetree{,time{,col{,scale{,prec}}}))
```

Here `onetree` is any matrix output of the quantlet ITT: `Ttree`, `P`, `Q`, `AD` or `LocVol`. `time` is an optional parameter, in default it is equidistant time axe, but it can be also equal to the output parameter `Time` of ITT. Parameter `col` is (2x1) vector, which specify the colours in the tree. The first row is the color of the mesh and the second row the colour of its description. `scale` is a boolean determining if the tree should be scaled by the values given in `onetree` or if a particular constant-volatility trinomial tree should be plotted and the values of `onetree` will be used only for description.

This is, for instance, used in the quantlet `plotITT` for plotting the transition probabilities whose values have nothing common with the vertical spacing in the tree. `prec` is a scalar specifying the precision of the description. The output parameter `nodes` is a graphical object again.

The last graphical tool is the quantlet `grITTstsp`:

```
{nodes,probs,axe,on,op} = grITTstsp (itree{,text{,col{,prec}}}))
```

Variable `itree` is the output of ITT again, `text` and `prec` have the same meaning as before, but the parameter `col` has now five rows. They correspond in the following order to the colour of the tree mesh, of the nodes description, of the probabilities description, of the overwritten nodes, and lastly, of the colour of overwritten probabilities. Here we have five output parameters – all of them are composed graphical objects corresponding in sequence to the tree nodes, to the description of the probabilities, to the time axe, to the overwritten nodes and to the overwritten probabilities. For instance, Figure 3.15 was produced by the following commands (the ITT `impltree` has been already computed before):

■ STFitt02.xpl

Note that if there were no overwrites in the nodes or probabilities, `on` or `op` are equal to NaN and it is not possible to show them.


			123.50
		113.08	114.61
	105.22	106.34	113.60
100.00	100.00	100.01	106.82
	95.04	94.04	100.46
		88.43	87.27
			80.97

If one wants to get a text output (the graphical output may be too "wild" and its description unreadable), it is possible to write the result in the [XploRe](#) output window as a string:

```
ITTnicemat (title,onetree,prec)
```

Here `title` is a string with desired title of the matrix, `onetree` is any output (tree) matrix from the quantlet `ITT` and `prec` specifies the precision. You can compare Figure ??(a) with the result of `nicemat`: `ITTnicemat`

State Space Matrix

 STFitt05.xpl

Last in this section we will introduce a quantlet, which actually belongs into the next section, but it is also possible to use it separately. `ITTcrr` is called by `ITT` to compute the option price. It builds up a constant-volatility trinomial tree (combining two steps of a standard CRR binomial tree), its formal definition is:

```
{ttree,optprice} = ITTcrr (S, K, r, sigma, time, opt, div)
```

where `S` is the the spot price of the underlying, `K` is the exercise price, `r` is the continuous interest rate ($\in (0,1)$), `sigma` is the constant volatility, `time` is the vector of time points, `opt = 1` specifies a call option and `opt = 0` a put option. `div` is the dividend rate ($\in (0,1)$). The output parameters are `ttree` and `optprice`, the constant-volatility trinomial tree and the price of given option respectively.

3.5.3 What is hidden

This subsection comments on auxiliary quantlets which are not intended for a direct usage. First we present the quantlet `ITTnewnodes`, which computes a new level of an implied trinomial tree. It also checks that the forward condition from Section 3.4.4 is fulfilled.

```
{newnodes,overwritten} = ITTnewnodes (old, r, sigma, deltat, div)
```

`old` is a vector containing the underlying prices on last known level of the tree. `r` and `div` are the continuous interest rate and dividend yield, `sigma` is the volatility and `deltat` is the time step between the last and the new level $\Delta t = t_{n+1} - t_n$. The outputs are two vectors: the vector with the prices in the new level `newnodes` and the vector `overwritten`. This corresponds to `newnodes`, it is firstly filled with NaN's and afterward, if a node is overwritten, the corresponding node in `overwritten` contains its new value. It is worth to mention how the implausible price will be overwritten: If the highest new underlying price S_1 is lower than the forward price F_1 at the node $s_{n,1}$, it will be set up one percent higher than the forward, and similarly, any lowest new underlying price S_{2n+1} which is higher than the forward price F_{2n-1} of the node $s_{n,2n-1}$ will be set 1% lower than this forward price.

The Arrow-Debreu prices will be computed stepwise with the same philosophy, like the state space, with the difference that no overwrites are necessary.

```
newad = ITTad (up, down, ad, r, deltat)
```

`ITTad` computes the Arrow-Debreu prices of an Implied Trinomial Tree on one particular level. `up` and `down` are vectors of up and down transition probabilities going from the last level of known Arrow-Debreu prices `ad`. The new computed prices are then returned in the vector `newad`.

```
fx = ITTterm (t)
```

`ITTterm` represents the non-linear equation system (3.17) – it computes its value in time points given by the vector `t`. The equation system is solved with the quantlet `nmnewton`. The Jacobian matrix necessary for the numerical solution is returned by quantlet `ITTtermder`:

```
jac = ITTtermder (t)
```


Bibliography

- Derman, E. and Kani, I. (1994). The Volatility Smile and Its Implied Tree. *Quantitative Strategies Research*, <http://gs.com/qs>, Goldman Sachs.
- Derman, E., Kani, I. and Chriss, N. (1996). Implied Trinomial Trees of the Volatility Smile. *Quantitative Strategies Research*, <http://gs.com/qs>, Goldman Sachs.
- Härdle, W., Hlávka, Z. and Klinke, S. (2000). *XploRe Application Guide*. Springer Verlag, Heidelberg.
- Härdle, W., Kleinow, T. and Stahl, G. (2002). *Applied Quantitative Finance*. Springer-Verlag, Berlin.
- Hull, J (1989). *Options, Futures and Other Derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Komorád, K (2002). Implied Trinomial Trees and Their Implementation with XploRe. *Bachelor Thesis*, HU Berlin.
- Ross, S., Westerfield, R. and Jaffe, J. (2002). *Corporate Finance*. Mc Graw-Hill.

4 Functional data analysis

Michal Benko, Wolfgang Härdle

4.1 Introduction

5 Nonparametric Productivity Analysis

Wolfgang Härdle, Seok-Oh Jeong

5.1 Introduction

Productivity analysis is an important part of classical microeconomic theory. Farrell (1957) studied the problem of measuring efficiency in an empirical way. Shephard (1970) formed the foundation of the modern theory of cost and frontier function. Färe, Grosskopf, and Lovell (1985) and Färe, Grosskopf, and Lovell (1994) provide a comprehensive survey of the problem of measuring efficiency and the theoretical formulations of frontier function.

The activity of production units such as banks, universities, governments and hospitals may be described and formalized by *the production set* of feasible input-output points (x, y) :

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \text{ can produce } y\}.$$

In many cases y is one-dimensional, so Ψ can be characterized by a function g called the *frontier function* or the *production function*:

$$\Psi = \{(x, y) \in \mathbb{R}_+^{p+1} \mid y \leq g(x)\}.$$

The *input (requirement) set* $X(y)$ is defined by:

$$X(y) = \{x \in \mathbb{R}_+^p \mid (x, y) \in \Psi\},$$

which is the set of all input vectors $x \in \mathbb{R}_+^p$ that yield at least the output vector y . The *output (correspondence) set* $Y(x)$ defined by:

$$Y(x) = \{y \in \mathbb{R}_+^q \mid (x, y) \in \Psi\},$$

which is the set of all output vectors $y \in \mathbb{R}_+^q$ that is obtainable from the input vector x .

In productivity analysis one is interested in the input and output *isoquants* or *efficient boundaries*, denoted by $\partial X(y)$ and $\partial Y(x)$ respectively. They consist of the minimal attainable boundary in a radial sense: for $y \neq 0$,

$$\partial X(y) = \{x \mid x \in X(y), \theta x \notin X(y), 0 < \theta < 1\}$$

and for $Y(x) \neq \{0\}$,

$$\partial Y(x) = \{y \mid y \in Y(x), \lambda y \notin X(y), \lambda > 1\}.$$

For $y = 0$ we have $\partial X(y) = \{0\}$, and for $Y(x) = \{0\}$ $\partial Y(x) = \{0\}$.

Returns to scale is a characteristic of the surface of the production set. The production set exhibits *constant returns to scale* (CRS) if, for $\alpha \geq 0$ and $P \in \Psi$, $\alpha P \in \Psi$; it exhibits *non-increasing returns to scale* (NIRS) if, for $0 \leq \alpha \leq 1$ and $P \in \Psi$, $\alpha P \in \Psi$; it exhibits *non-decreasing returns to scale* (NDRS) if, for $\alpha \geq 1$ and $P \in \Psi$, $\alpha P \in \Psi$. Especially, a convex production set exhibits non-increasing returns to scale, but note that the converse is not true.

Recall that, given a production set Ψ with the scalar output y , the production function g can also be defined for $x \in \mathbb{R}_+^p$:

$$g(x) = \sup\{y \mid (x, y) \in \Psi\}.$$

It may be defined via the input set and the output set as well:

$$g(x) = \sup\{y \mid x \in X(y)\} = \sup\{y \mid y \in Y(x)\}.$$

We can interpret returns to scale in terms of the production function g :

- Constant returns to scale: For all $t \geq 0$, $g(tx) = tg(x)$.
- Non-decreasing returns to scale: For all $t \geq 1$, $g(tx) \geq tg(x)$
- Non-increasing returns to scale: For all $t \geq 1$, $g(tx) \leq tg(x)$.

How can one evaluate the performance of a given pair of input and output? When the outputs are scalars, one may do it with input x and output y by computing $g(x) - y$ or $y/g(x)$. This is usually inadequate though when multiple inputs or outputs are incorporated. In order to tackle this problem *radial*

efficiency measures have been proposed. For a given input-output point (x_0, y_0) , its input efficiency is defined as

$$\theta^{\text{IN}}(x_0, y_0) = \inf\{\theta \mid \theta x_0 \in X(y_0)\}.$$

The efficient level of input corresponding to the output level y_0 is then given by

$$x^\partial(y_0) = \theta^{\text{IN}}(x_0, y_0)x_0.$$

Note that $x^\partial(y_0)$ is the intersection of $\partial X(y_0)$ and the ray θx_0 , $\theta > 0$. The output efficiency score $\theta^{\text{OUT}}(x_0, y_0)$ can be defined similarly:

$$\theta^{\text{OUT}}(x_0, y_0) = \sup\{\theta \mid \theta y_0 \in Y(x_0)\}.$$

The efficient level of output corresponding to the input level x_0 is given by

$$y^\partial(x_0) = \theta^{\text{OUT}}(x_0, y_0)y_0.$$

5.2 Nonparametric Hull Methods

5.2.1 An Overview

The production set Ψ and the production function g is usually unknown, but typically a sample of production units or decision making units (DMU's) is available instead:

$$\mathcal{X} = \{(x_i, y_i), i = 1, \dots, n\}.$$

The problem of productivity analysis is to estimate Ψ or g from the data \mathcal{X} .

The most popular nonparametric method is based on data envelopment analysis (DEA), which is an extension of Farrel (1957)'s idea and was popularized in a modern operational research fashion by Charnes, Cooper, and Rhodes (1978). Deprins, Simar, and Tulkens (1984), who extended the DEA to the more flexible class of production sets, introduced the free disposal hull (FDH) model dropping convexity assumption for the production set.

Statistical properties of these hull methods such as consistency and minimax rate of convergence have been studied in the literature. Park (2001) and Simar and Wilson (1999) provide insightful reviews on the statistical inference of existing nonparametric frontier models.

5.2.2 Data Envelopment Analysis

The data envelopment analysis (DEA) of the observed sample \mathcal{X} is defined as the smallest free disposable and convex set containing \mathcal{X} :

$$\widehat{\Psi}_{\text{DEA}} = \left\{ (x, y) \in \mathbb{R}_+^{p+q} \mid x \geq \sum_{i=1}^n \gamma_i x_i, y \leq \sum_{i=1}^n \gamma_i y_i \right. \\ \left. \text{for some } (\gamma_1, \dots, \gamma_n) \text{ such that } \sum_{i=1}^n \gamma_i = 1, \gamma_i \geq 0 \forall i \right\},$$

where the inequalities between vectors are understood componentwise. A set Ψ is *free disposable*, if for given $(x, y) \in \Psi$ all (x', y') with $x' \geq x$ and $y' \leq y$ belong to Ψ . Free disposability is generally assumed in economic modeling. The efficiency scores for a given input-output level (x_0, y_0) are estimated by :

$$\widehat{\theta}^{\text{IN}}(x_0, y_0) = \min\{\theta > 0 \mid (\theta x_0, y_0) \in \widehat{\Psi}_{\text{DEA}}\}; \\ \widehat{\theta}^{\text{OUT}}(x_0, y_0) = \max\{\theta > 0 \mid (x_0, \theta y_0) \in \widehat{\Psi}_{\text{DEA}}\}.$$

The efficient levels for a given level (x_0, y_0) are estimated by

$$\widehat{x}^\partial(y_0) = \widehat{\theta}^{\text{IN}}(x_0, y_0)x_0; \quad \widehat{y}^\partial(x_0) = \widehat{\theta}^{\text{OUT}}(x_0, y_0)y_0.$$

The consistency and the convergence rate of this radial efficiency score with multidimensional inputs and outputs were established analytically by Kneip, Park, and Simar (1998). For $p = 1$ and $q = 1$, Gijbels, Mammen, Park, and Simar (1999) obtained its limit distribution depending on the curvature of the frontier and the density at boundary. Jeong and Park (2002) extended this result to higher dimension.

5.2.3 Free Disposal Hull

The free disposal hull (FDH) of the observed sample \mathcal{X} is defined as the smallest free disposable set containing \mathcal{X} :

$$\widehat{\Psi}_{\text{FDH}} = \{(x, y) \in \mathbb{R}_+^{p+q} \mid x \geq x_i, y \leq y_i, i = 1, \dots, n\}.$$

One define the estimators of efficiency scores by FDH in a similar way to DEA. By a straightforward calculation we obtain their closed forms for a given input-

output level (x_0, y_0) :

$$\widehat{\theta}^{\text{IN}}(x_0, y_0) = \min_{i: y \leq y_i} \max_{1 \leq j \leq p} \frac{x_i^j}{x_0^j};$$

$$\widehat{\theta}^{\text{OUT}}(x_0, y_0) = \max_{i: x \geq x_i} \min_{1 \leq k \leq q} \frac{y_i^k}{y_0^k},$$

where v^j is the j th component of a vector v . The efficient levels for a given level (x_0, y_0) are estimated by the same way as those for DEA. Park, Simar, and Weiner (1999) showed that the limit distribution of the FDH estimator with multivariate setup is a Weibull distribution depending upon the slope of the frontier and the density at boundary.

5.3 DEA in Practice : Insurance Agencies

The quantlet `deahull` for DEA has the following syntax :

```
dea = deahull (X, Y)
  computes the input and output efficiency score for each DMU
  based on DEA
```

with input variables: X is a $n \times p$ matrix, p dimensional inputs of n DMU's, and Y is $n \times q$ matrix, q dimensional outputs of n DMU's.

The list `dea` contains the output variables : `dea. effscore` is a n dimensional vector containing the input efficiency scores, $\widehat{\theta}^{\text{IN}}$, of n DMU's, and `dea. efflevel` is $n \times p$ matrix containing the estimates of efficient level of inputs, \widehat{x}^ϑ , of n DMU's.

To illustrate how DEA works, consider an example from the empirical study by Scheel (1999). He studied the efficiency of 63 agencies of a German insurance company by the DEA method. The input ($X \in \mathbb{R}_+^4$) and output ($Y \in \mathbb{R}_+^2$) variables were as follows:

X_1 : Number of clients of Type A

X_2 : Number of clients of Type B

Table 5.1: Summary Statistics for 63 agencies of a German insurance company

	Minimum	Maximum	Mean	Median	Std.Error
X1	42	572	225.54	197	131.73
X2	55	481	184.44	141	110.28
X3	0	140	19.762	10	26.012
X4	73756	693820	258670	206170	160150
Y1	2	70	22.762	16	16.608
Y2	696	33075	7886.7	6038	7208

X_3 : Number of clients of Type C

X_4 : Potential new premiums in EURO

Y_1 : Number of new contracts

Y_2 : Sum of new premiums in EURO

The two output variables are typical for insurance agencies. Summary statistics for this data are given in Table 5.1. The included [XploRe](#) code reads the data `agency.dat`, creates the input variables for the quantlet `deahull`, runs the quantlet, and lists the efficiency and the efficient level of inputs for each agency as a result, see Table 5.2 and Table 5.3.

The input efficient scores `effscore` are useful as a benchmark in comparing the 63 DMU's, the insurance agencies. The list of efficient level of inputs provides with a 'goal' inputs for each agency to be efficient.

5.4 FDH in Practice : Manufacturing Industry

The quantlet `fdhull` for FDH has the following syntax:

```
eff = fdhull (X, Y)
  computes the input and output efficiency score for each DMU
  based on FDH
```


Table 5.2: Efficiency score of the 63 DMU's insurance Agencies

Effeciency score	
1	0.38392
2	0.49063
3	0.86449
.	.
.	.
.	.
62	0.79892
63	1

STFagency.xpl

Table 5.3: Efficiency level of the 63 DMU's insurance Agencies

efflevel				
1	52.981	92.909	3.8392	108960
2	81.444	60.838	2.4531	76895
3	131.4	72.617	2.5935	96070
.
.
.
62	66.311	87.083	1.5978	111710
63	108	257	0	299910

STFagency.xpl

with input variables: \mathbf{X} is $n \times p$ matrix, p dimensional inputs of n DMU's, and \mathbf{Y} is $n \times q$ matrix, q dimensional outputs of n DMU's.

The arbitrary name `eff` is used to indicate the output variable: `eff` which is the $n \times 2$ matrix containing the input and output efficiency scores of n DMU's.

In order to illustrate how this quantlet works the Manufacturing Industry Productivity Database from the National Bureau of Economic Research, USA is considered. This database is available in the internet [<http://www.nber.org>]

Table 5.4: Summary Statistics for Manufacturing Industry Productivity Database (NBER, USA)

	Minimum	Maximum	Mean	Median	Std.Error
X1	0.8	500.5	37.833	21	54.929
X2	18.5	145130	4313	1957.2	10771
X3	0.5	3807.8	139.96	49.7	362
X4	15.8	64590	2962.8	1234.7	6271.1
Y	34.1	56311	3820.2	1858.5	6392

with a description of the database. It contains annual industry-level data on output, employment, payroll and other input costs, investment, capital stocks, and various industry-specific price indexes from 1958 on hundreds of manufacturing industries (indexed by 4 digits numbers) in the United States. We chose the data from 1996 (458 DMU's), for example, with 4 input variables ($p = 4$) and 1 output variable ($q = 1$) along with the study of Park, Simar, and Weiner (1999) :

X_1 : Total employment

X_2 : Total cost of materials

X_3 : Cost of electric and fuels

X_4 : Total real capital stock

Y : Total value added


Summary statistics for this data are given in Table 5.4. The included [XploRe](#) code reads the MS-Excel data file `nber96.csv`, creates the vector of the ID for DMU's and the input variables for the quantlet `fdhull`, runs the quantlet, and lists the efficiency scores, $\hat{\theta}^{\text{IN}}(\cdot, \cdot)$ and $1/\hat{\theta}^{\text{OUT}}(\cdot, \cdot)$, with ID of industries.

From Table 5.5, we see that the DMU indexed by 2015 is efficient in both input and output oriented aspect. This means the DMU is one of the vertices of the free disposal hull generated by the 458 observations. On the other hand, the DMU 2298 seems to be good in input efficiency but poor in output efficiency.

With these efficiency scores we can estimate the efficient level of input (or output) by multiplying (or dividing) the score to the corresponding observation.

Table 5.5: FDH efficiency scores of the 458 observations

fdh			
1	2011	0.93803	0.97451
2	2013	0.64321	0.70255
3	2015	1	1
.	.	.	.
.	.	.	.
.	.	.	.
75	2298	0.95588	0.47239
.	.	.	.
.	.	.	.
.	.	.	.
458	3999	0.98961	0.94792

 STFnber96.xpl

For example, DMU 2013 had inputs $X_1 = 88.1$, $X_2 = 14925$, $X_3 = 250$, $X_4 = 4365.1$, and output $Y = 5954.2$ with input efficiency 0.64321 and with output efficiency 0.70255. Then the efficient level of input for this firm is given by $X_1 = 56.667$, $X_2 = 9600$, $X_3 = 160.8$, and $X_4 = 2807.7$. And in the aspect of output efficiency this firm should have increased their output by $Y = 4183.1$ with the observed level of inputs to be considered as output technically efficient.

Bibliography

- Charnes, A., Cooper, W. W, and Rhodes, E. (1978). Measuring the Inefficiency of Decision Making Units, *European Journal of Operational Research* **2**, 429–444.
- Deprins, D., Simar, L. and Tulkens, H. (1984). Measuring Labor Inefficiency in Post Offices, in *The Performance of Public Enterprises: Concepts and Measurements*, 243–267.
- Färe, R., Grosskopf, S. and Lovell, C. A. K. (1985). *The Measurement of Efficiency of Production*, Kluwer-Nijhoff.
- Färe, R., Grosskopf, S. and Lovell, C. A. K. (1994). *Production Frontiers*, Cambridge University Press.
- Farrell, M. J. (1957). The Measurement of Productivity Efficiency, *Journal of the Royal Statistical Society, Ser. A* **120**, 253–281.
- Gijbels, I., Mammen, E., Park, B. U. and Simar, L. (1999). On Estimation of Monotone and Concave Frontier Functions, *Journal of the American Statistical Association* **94**, 220–228.
- Jeong, S. and Park, B. U. (2002). Limit Distributions Convex Hull Estimators of Boundaries, submitted.
- Kneip, A., Park, B. U. and Simar, L. (1998). A Note on the Convergence of Nonparametric DEA Efficiency Measures, *Econometric Theory* **14**, 783–793.
- Park, B. U. (2001). On Nonparametric Estimation of Data Edges, *Journal of the Korean Statistical Society* **30**, 2, 265–280.
- Park, B. U., Simar, L. and Weiner, Ch. (1999). The FDH Estimator for Productivity Efficiency Scores : Asymptotic Properties, *Econometric Theory* **16**, 855–877.

-
- Scheel, H. (1999). Continuity of the BCC efficiency measure, in: Westermann (ed.), *Data Envelopment Analysis in the public and private service sector*, Gabler, Wiesbaden.
- Shephard, R. W. (1970). *Theory of Cost and Production Function*, Princeton University Press.
- Simar, L. and Wilson, P. W. (1999). Statistical Inference in Nonparametric Frontier Models : The State of the Art, Discussion Paper # 9904, Institute de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

6 Money Demand Modelling

Noer Azam Achsani, Oliver Holtemöller and Hizir Sofyan

6.1 Introduction

Money demand is one of the most important indicators for the monetary situation in a country. The relationship between money balances and their determinants has been studied extensively. From the previous researches, there is an overall agreement on the view that the knowledge of the demand for money function is important for the formulation of an appropriate monetary policy.

In this paper, we explore the M2 money demand function for Indonesia in the more recent period 1990:I - 2002:III. This period is dominated by the Asian crises, which started in 1997. In the aftermath of the crisis, a number of immense financial and economic problems have emerged in Indonesia. The state has deteriorated, and inflation rates have risen in the middle of 1997 until the middle of 1998. The price level increased by about 16 percent in 1997 compared to the previous year. In the same period, the call money rate increased temporarily from 12.85 percent to 57.10 percent and the money stock increased by about 54 percent. The economy contracted as reflected by the decrease in the GNP of about 11 percent. Given these extraordinary economic developments, it may not be expected that a stable money demand function existed during that period.

This research is at least important in two different respects. Firstly, it explores money demand using data of Indonesia, a very dynamic but relatively unexplored country. Secondly, we employ not only standard econometric model (which are normally used in money demand), but also a relatively new approach in economics analysis, so called fuzzy Takagi-Sugeno model.

The fuzzy model can be used to explain the underlying structure of the systems. In the real world, it is almost impossible to find a perfect situation in which

all variables are available and countable. There are always some unpredictable factors which influence the system. The use of fuzzy Takagi-Sugeno model as an alternative may give us a complementary solution to the common one.

The rest of the paper is organized as follows. In the Section 2 we introduce the money demand function and estimate the model. Section 3 performs the fuzzy approach and its application to money demand. Last section presents the conclusions of this research.

6.2 Money Demand

6.2.1 General Remarks and Literature

Major central banks stress the importance of money growth analysis and of stable money demand function for monetary policy purposes. The Deutsche Bundesbank, for example, has followed an explicit monetary targeting strategy from 1975 to 1998, and the analysis of monetary aggregates is one of the two pillars of the European Central Bank's (ECB) monetary policy strategy. Details about these central banks' monetary policy strategies, a comparison and further references can be found in Holtemöller (2002). The research on the existence and stability of a money demand function is motivated inter alia by the following two observations: (i) Money growth is highly correlated with inflation, see McCandless and Weber (1995) for international empirical evidence. Therefore, monetary policy makers use money growth as one indicator for future risks to price stability. The information content of monetary aggregates for future inflation assessment is based on a stable relationship between money, prices and other observable macroeconomic variables. This relationship is usually analyzed in money demand studies. (ii) The monetary policy transmission process is still a "black box", see Mishkin (1995) and Bernanke and Gertler (1995). If we are able to specify a stable money demand function, an important element of the monetary transmission mechanism is revealed and may help to learn more about monetary policy transmission.

A huge literature exists about the analysis of money demand. The majority of the studies is concerned with industrial countries. Examples are Hafer and Jansen (1991), Miller (1991), McNown and Wallace (1992) and Mehra (1993) for the USA; Lütkepohl and Wolters (1999), Coenen and Vega (1999), Brand and Cassola (2000) and Holtemöller (2004) for the Euro area; Arize and Shwiff (1993), Miyao (1996) and Bahmani-Oskooee (2001) for Japan; Drake

and Chrystal (1994) for the UK; Haug and Lucas (1996) for Canada; Lim (1993) for Australia and Orden and Fisher (1993) for New Zealand.

There is also a growing number of studies analyzing money demand in developing and emerging countries, primarily triggered by the concern among central bankers and researchers around the world about the impact of moving toward flexible exchange rate regimes, globalization of capital markets, ongoing financial liberalization, innovation in domestic markets, and the country-specific events on the demand for money (Sriram, 1999). Examples are Hafer and Kutan (1994) and Tseng *et. al* (1994) for China; Moosa (1992) for India; Arize (1994) for Singapore and Deckle and Pradhan (1997) for ASEAN countries.

For Indonesia, a couple of studies have been conducted by applying the error correction framework. The results, however, seem to be contradictory. Price and Insukindro (1994) use quarterly data from 1969:I to 1987:IV. The results were based on three different methods of testing for cointegration. Eagle Grager method show that there was weak evidence of cointegrating relationship. Furthermore Johansen's cointegration technique found up to two cointegrating vectors, but the error correction model (ECM) didn't find a significant relationship. In contrast, Deckle and Pradhan (1997), who use annual data, do not find any cointegrating relationship that can be interpreted as money demand function.

6.2.2 Econometric Specification of Money Demand Functions

The starting point of empirical money demand analysis is the choice of variables to be included in the money demand function. It is common practice to assume that the desired level of nominal money demand depends on the price level, a transaction (or scaling) variable, and a vector of opportunity costs, see for example Goldfeld and Sichel (1990), Ericsson (1999) and Holtemöller (2003):

$$(M^*/P) = f(Y, R_1, R_2, \dots), \quad (6.1)$$

where M^* is nominal money demanded, P is the price level, Y is real income (the transaction variable), and R_i are the elements of the vector of opportunity costs which possibly also includes the inflation rate. A money demand function of this type is not only the result of traditional money demand theories but also of modern micro-founded dynamic stochastic general equilibrium models, see

for example Walsh (1998). An empirical standard specification of the money demand function is the partial adjustment model (PAM). (Goldfeld and Sichel, 1990) show that a desired level of real money holdings $MR_t^* = M_t^*/P_t$:

$$\ln MR_t^* = \phi_0 + \phi_1 \ln Y_t + \phi_2 R_t + \phi_3 \pi_t, \quad (6.2)$$

where R_t represents one or more interest rates and $\pi_t = \ln(P_t/P_{t-1})$ is the inflation rate, and an adjustment cost function

$$C = \alpha_1 [\ln M_t^* - \ln M_t]^2 + \alpha_2 [(\ln M_t - \ln M_{t-1}) + \delta (\ln P_t - \ln P_{t-1})]^2 \quad (6.3)$$

yield the following reduced form

$$\ln MR_t = \mu\phi_0 + \mu\phi_1 \ln Y_t + \mu\phi_2 R_t + (1 - \mu) \ln MR_{t-1} + \gamma\pi_t, \quad (6.4)$$

where

$$\mu = \alpha_1 / (\alpha_1 + \alpha_2) \quad \text{and} \quad \gamma = \mu\phi_3 + (1 - \mu)(\delta - 1). \quad (6.5)$$

The parameter δ controls whether nominal money ($\delta = 0$) or real money ($\delta = -1$) adjusts. Intermediate cases are also possible. Notice that the coefficient to the inflation rate depends on the value of ϕ_3 and on the parameters of the adjustment cost function. The imposition of price-homogeneity is rationalized by economic theory and Goldfeld and Sichel (1990) proposed that empirical rejection of the unity of the price level coefficient should be interpreted as an indicator for misspecification. The reduced form can also be augmented by lagged independent and further lagged dependent variables in order to allow for a more general adjustment process.

Rearranging (6.4) yields:

$$\begin{aligned} \Delta \ln MR_t &= \mu\phi_0 + \mu\phi_1 \Delta \ln Y_t + \mu\phi_1 \ln Y_{t-1} + \mu\phi_2 \Delta R_t \\ &\quad + \mu\phi_2 R_{t-1} - \mu \ln MR_{t-1} + \gamma \Delta \pi_t + \gamma \pi_{t-1} \\ &= \mu\phi_0 - \mu \left(\ln MR_{t-1} - \phi_1 \ln Y_{t-1} - \phi_2 R_{t-1} - \frac{\gamma}{\mu} \pi_{t-1} \right) \\ &\quad + \mu\phi_1 \Delta \ln Y_t + \mu\phi_2 \Delta R_t + \gamma \Delta \pi_t. \end{aligned} \quad (6.6)$$

Accordingly, the PAM can also be represented by an error-correction model like (6.6).

Since the seminal works of Nelson and Plosser (1982), who have shown that relevant macroeconomic variables exhibit stochastic trends and are only stationary after differencing, and Engle and Granger (1987), who introduced the concept of cointegration, the (vector) error correction model, (V)ECM, is the dominant econometric framework for money demand analysis. If a certain set of conditions about the number of cointegration relations and exogeneity properties is met, the following single equation error correction model (SE-ECM) can be used to estimate money demand functions:

$$\begin{aligned} \Delta \ln MR_t = & c_t + \alpha \underbrace{(\ln MR_{t-1} - \beta_2 \ln Y_{t-1} - \beta_3 R_{t-1} - \beta_4 \pi_{t-1})}_{\text{error correction term}} \\ & + \sum_{i=1}^k \gamma_{1i} \Delta \ln MR_{t-i} + \sum_{i=0}^k \gamma_{2i} \Delta \ln Y_{t-i} + \sum_{i=0}^k \gamma_{3i} \Delta R_{t-i} + \sum_{i=0}^k \gamma_{4i} \Delta \pi_{t-i}, \end{aligned} \quad (6.7)$$

It can be shown that (6.4) is a special case of the error correction model (6.7). In other words, the PAM corresponds to a SE-ECM with certain parameter restrictions. The SE-ECM can be interpreted as a partial adjustment model with β_2 as long-run income elasticity of money demand, β_3 as long-run semi-interest rate elasticity of money demand, and less restrictive short-run dynamics. The coefficient β_4 , however, cannot be interpreted directly.

In practice, the number of cointegration relations and the exogeneity of certain variables cannot be considered as known. Therefore, the VECM is the standard framework for empirical money demand analysis. In this framework, all variables are assumed to be endogenous a priori, and the imposition of a certain cointegration rank can be justified by statistical tests. The standard VECM is obtained from a vector autoregressive (VAR) model

$$x_t = \mu_t + \sum_{i=1}^k A_i x_{t-i} + u_t, \quad (6.8)$$

where x_t is a $(n \times 1)$ -dimensional vector of endogenous variables, μ_t contains deterministic terms like constant and time trend, A_i are $(n \times n)$ -dimensional

coefficient matrices and $u_t \sim N(0, \Sigma_u)$ is a serially uncorrelated error term. Subtracting x_{t-1} and rearranging terms yields the VECM

$$\Delta x_{t-1} = \mu_t + \Pi x_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta x_{t-i} + u_t. \quad (6.9)$$

Π and Γ_i are functions of the A_i . The matrix Π can be decomposed into two $(n \times r)$ -dimensional matrices α and β : $\Pi = \alpha\beta'$ where α is called adjustment matrix, β comprises the cointegration vectors, and r is the number of linearly independent cointegration vectors (cointegration rank). Following Engle and Granger (1987), a variable is integrated of order d , or $I(d)$, if it has to be differenced d -times to become stationary. A vector x_t is integrated of order d if the maximum order of integration of the variables in x_t is d . A vector x_t is cointegrated, or $CI(d, b)$, if there exists a linear combination $\beta'x_t$ that is integrated of a lower order $(d - b)$ than x_t .

6.2.3 Estimation of Indonesian Money Demand

We use quarterly data from 1990:1 until 2002:3 for our empirical analysis. The data is not seasonally adjusted and taken from Datastream (gross national product at 1993 prices Y and long-term interest rate R) and from Bank Indonesia (money stock M2 M and consumer price index P). In the following, logarithms of the respective variables are indicated by small letters, and $mr = \ln M - \ln P$ denotes logarithmic real balances. The data is depicted in figure 6.1.

Before the appropriate econometric methodology can be chosen, the stochastic properties of the variables have to be analyzed. Table 6.1 presents the results of unit root tests for logarithmic real balances mr , logarithmic real GNP y , logarithmic price level p , and logarithmic long-term interest rate r .

The table 6.1 shows unit root test results for the variables indicated in the first column. The second column describes deterministic terms included in the test regression: constant (c), seasonal dummies (s), linear trend (t), and shift and impulse dummies according to the model c in Perron (1989) (P89c) allowing for a change in mean and in the slope of a linear trend. The break point is given in parentheses. Lags denotes the number of lags included in the test regression. Column CV contains critical values. Three (two) asterisks denote significance at the 1% (5%) level.

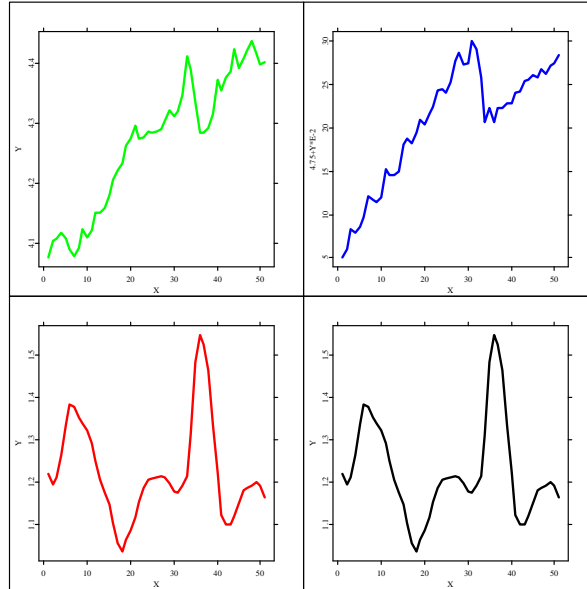


Figure 6.1: Data

Table 6.1: Unit Root Tests

Variable	Deterministic Terms	Lags	Test Statistic	1/5/10% CV
<i>mr</i>	c, t, s, P89c (98:3)	2	-4.55 **	-4.75 / -4.44 / -4.18
<i>y</i>	c, t, s, P89c (98:1)	0	-9.40 ***	-4.75 / -4.44 / -4.18
<i>p</i>	c, t, s, P89c (98:1)	2	-9.46 ***	-4.75 / -4.44 / -4.18
<i>r</i>	c, s	2	-4.72 ***	-3.57 / -2.92 / -2.60

Because the time series graphs show that there seem to be structural breaks in real money, GNP and price level, we allow for the possibility of mean shift and change in the slope of a linear trend in the augmented Dickey-Fuller test regression. This corresponds to model (c) in Perron (1989), where the critical values for this type of test are tabulated. In the unit root test for the interest

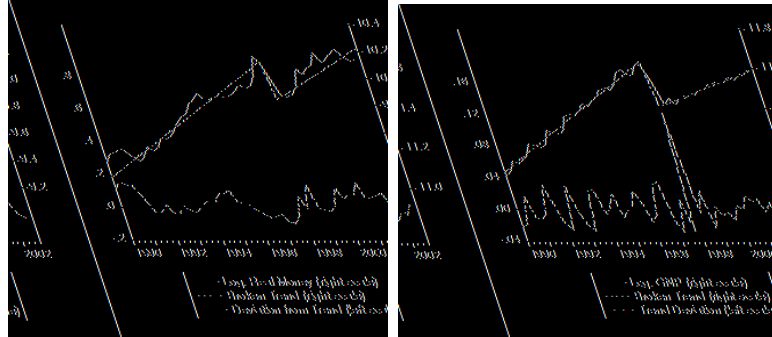


Figure 6.2: Fitted Trends for Real Money and Real GNP

rate, only a constant is considered. According to the test results, real money, real GNP and price level are trend-stationary, that is they do not exhibit a unit root, and the interest rate is also stationary. These results are quite stable with respect to the lag length specification. The result of trend-stationarity is also supported by visual inspection of a fitted trend and the corresponding trend deviations, see figure 6.2. In the case of real money, the change in the slope of the linear trend is not significant.

Indonesian money demand is now estimated by OLS using the reduced form equation 6.4 (s_{it} are centered seasonal dummies, ds denotes a step dummy switching from zero to one in the respective quarter, di is an impulse dummy having value one only in the respective quarter, t -values in parantheses):

$$\begin{aligned}
 mr_t = & 0.531 mr_{t-1} + 0.470 y_t - 0.127 r_t \\
 & (6.79) \qquad (4.87) \qquad (-6.15) \\
 & - 0.438 - 0.029 s_{1t} - 0.034 s_{2t} - 0.036 s_{3t} \\
 & (-0.84) \quad (-2.11) \quad (-2.57) \quad (-2.77) \\
 & + 0.174 ds9802_t + 0.217 di9801_t + 0.112 ds9803_t + u_t(6.10) \\
 & (3.54) \qquad (5.98) \qquad (5.02)
 \end{aligned}$$

$$\begin{aligned}
 T = 50 (1990:1 - 2002:3) \quad R^2 = 0.987 \quad \text{RESET}(1) = 0.006 [0.941] \\
 \text{LM}(4) = 0.479 [0.751] \quad \text{JB} = 0.196 [0.906] \quad \text{ARCH}(4) = 0.970 [0.434]
 \end{aligned}$$

Here JB refers to the Jarque-Bera Test for nonnormality, RESET is the usual test for general nonlinearity and misspecification, LM(4) denotes a Lagrange-

Multiplier test for autocorrelation up to order 4, ARCH(4) is a Lagrange-Multiplier test for autoregressive conditional heteroskedasticity up to order 4. Numbers in brackets are corresponding p -values. Given these diagnostic statistics, the regression seems to be well specified. There is a mean shift in 98:3 and the impulse dummies capture the fact, that the structural change in GNP occurs two months before the change in real money. The inflation rate is not significant and is therefore not included in the equation.

The implied income elasticity of money demand is $0.47/(1-0.53) = 1$ and the interest rate elasticity is $-0.13/(1-0.53) = -0.28$. These are quite reasonable magnitudes. Equation (6.10) can be transformed into the following error correction representation:

$$\begin{aligned} \Delta mr_t = & -0.47 \cdot (mr_{t-1} - y_{t-1} + 0.28r_{t-1}) \\ & + 0.47\Delta y_t - 0.13\Delta r_t + \text{deterministic terms} + u_t \end{aligned} \quad (6.11)$$

Stability tests for the real money demand equation (6.10) are depicted in figure 6.3. The Cusum of squares test indicates some instability at the time of the Asian crises, and the coefficients of lagged real money and GNP seem to change slightly after the crisis. A possibility to allow for a change in these from 1998 on is to introduce two additional right-hand-side variables: lagged real money multiplied by the step dummy $ds9803$ and GNP multiplied by $ds9803$. The respective coefficients for the period 1998:3-2002:3 can be obtained by summing the coefficients of lagged real money and lagged real money times step dummy and of GNP and GNP times step dummy, respectively. This reveals that the income elasticity stays approximately constant ($0.28/(1-0.70)=0.93$ until 98:02 and $(0.28+0.39)/(1-0.70+0.32)=0.92$) from 98:3 to 2002:3 and that the interest rate elasticity declines in the second half of the sample from $-0.13/(1-0.70)=-0.43$ to $-0.13/(1-0.79+0.32)=-0.21$:

$$\begin{aligned} mr_t = & 0.697 \, mr_{t-1} + 0.281 \, y_t - 0.133 \, r_t \\ & (7.09) \qquad (2.39) \qquad (-6.81) \\ & - 0.322 \, mr_{t-1} \cdot ds9803_t + 0.288 \, y_t \cdot ds9803_t \\ & (-2.54) \qquad (2.63) \\ & + 0.133 - 0.032 \, s_{1t} - 0.041 \, s_{2t} - 0.034 \, s_{3t} \\ & (0.25) \qquad (-2.49) \qquad (-3.18) \qquad (-2.76) \\ & + 0.110 \, ds9802_t + 0.194 \, di9801_t + u_t \end{aligned} \quad (6.12)$$

(2.04) (5.50)

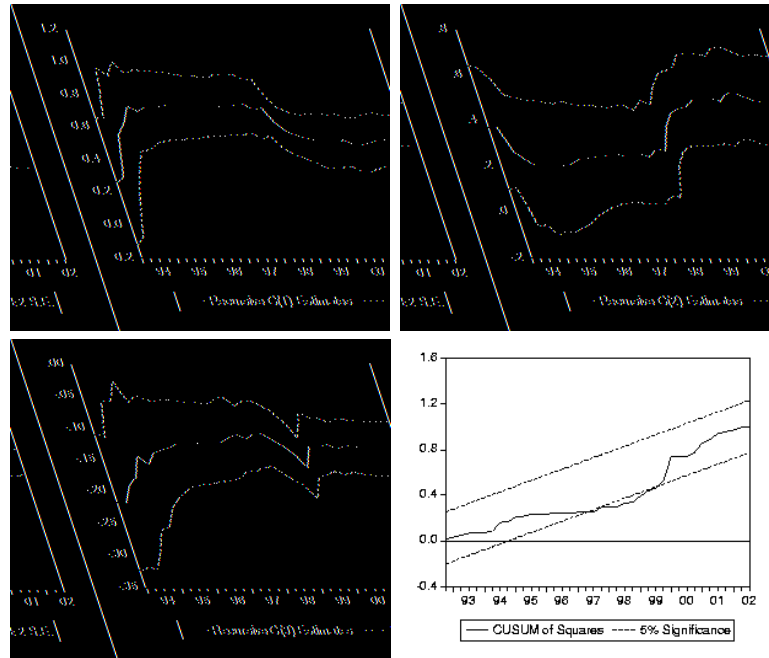


Figure 6.3: Stability Test for the Real Money Demand Equation (6.10)

$$\begin{array}{lll}
 T = 51 \text{ (1990:1 - 2002:3)} & R^2 = 0.989 & \text{RESET}(1) = 4.108 [0.050] \\
 \text{LM}(4) = 0.619 [0.652] & \text{JB} = 0.428 [0.807] & \text{ARCH}(4) = 0.408 [0.802]
 \end{array}$$

Accordingly, the absolute adjustment coefficient μ in the error correction representation increases from 0.30 to 0.62.

It can be concluded that Indonesian money demand has been surprisingly stable throughout and after the Asian crisis. A shift in the constant term and two impulse dummies that correct for the different break points in real money and real output are sufficient to yield a relatively stable money demand function with an income elasticity of one and an interest rate elasticity of -0.28 . However, a more flexible specification shows that the adjustment coefficient μ increases and that the interest rate elasticity decreases after the Asian crisis.

6.3 Fuzzy Model Identification

6.3.1 Fuzzy Clustering

Ruspini (1969) introduced a notion of fuzzy partition to describe the cluster structure of a data set and suggested an algorithm to compute the optimum fuzzy partition. Dunn (1973) generalized the minimum-variance clustering procedure to a Fuzzy ISODATA clustering technique. Bezdek (1981) generalized Dunn's approach to obtain an infinite family of algorithms known as the Fuzzy C-Means (FCM) algorithm. He generalized the fuzzy objective function by introducing the weighting exponent m , $1 \leq m < \infty$;

$$J_m(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d^2(x_k, v_i), \quad (6.13)$$

where $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^p$ is a subset of the real p -dimensional vector space \mathbb{R}^p consisted of n observations, U is a randomly fuzzy partition matrix of X into c parts, v_i is the cluster centers in \mathbb{R}^p , $d(x_k, v_i) = \|x_k - v_i\| = \sqrt{(x_k - v_i)^T (x_k - v_i)}$ is an inner product induced norm on \mathbb{R}^p , u_{ik} is referred to as the grade of membership or belonging of x_k to the cluster i . This grade of membership satisfies the following constraints:

$$0 \leq u_{ik} \leq 1, \quad \text{for } 1 \leq i \leq c, 1 \leq k \leq n, \quad (6.14)$$

$$0 < \sum_{k=1}^n u_{ik} < n, \quad \text{for } 1 \leq i \leq c, \quad (6.15)$$

$$\sum_{i=1}^c u_{ik} = 1, \quad \text{for } 1 \leq k \leq n. \quad (6.16)$$

The FCM uses an iterative optimization of the objective function, based on the weighted similarity measure between x_k and the cluster center v_i .

More on the steps of the FCM algorithm are discussed in Mucha and Sofyan (2000).

In practical applications, a validation method to measure the quality of a clustering result is needed. Its quality depends on many factors, such as the method of initialization, the choice of the number of classes c , and the clustering

method. The method of initialization requires a good estimate of the clusters and its application is dependent, so the cluster validity problem is reduced to the choice of an optimal number of classes c . Several cluster validity measures have been developed in the past by Bezdek and Pal (1992).

6.3.2 Takagi-Sugeno Approach

Takagi and Sugeno (1985) formulated the membership function of a fuzzy set A as $A(x), x \in X$. All the fuzzy sets are associated with linear membership functions. Thus, a membership function is characterized by two parameters giving the greatest grade 1 and the least grade 0. The truth value of a proposition "x is A and y is B " is expressed by

$$\|x \text{ is } A \text{ and } y \text{ is } B\| = A(x) \wedge B(y)$$

Based on above definition, the *affine* Takagi-Sugeno (TS) fuzzy model consists of rules R_i with the following structure:

If

$$x \text{ is } A_i \tag{6.17}$$

Then

$$y_i = a_i^\top x + b_i, \quad i = 1, 2, \dots, r. \tag{6.18}$$

This structure consists of two parts, namely x is A_i as *antecedent* part and $y_i = a_i^\top x + b_i$ as *consequent* part.

Note : $x \in \mathcal{X} \subset \mathbb{R}^p$ is a crisp input vector. A_i is a (multidimensional) fuzzy set: $\mu_{A_i}(x) : \mathcal{X} \rightarrow [0, 1]$, $y_i \in \mathbb{R}$ is an output of the i -th rule, $a_i \in \mathbb{R}^p$ is a parameter vector, b_i is a scalar. The index $i = 1, \dots, r$ shows that a given variable is related to i -th rule. r is the number of rules in the rule base.

Given the output of the individual *consequents* y_i , the global output y of the Takagi-Sugeno model is computed by using the fuzzy mean formula:

$$y = \frac{\sum_{i=1}^r \beta_i(x) y_i}{\sum_{i=1}^r \beta_i(x)} \tag{6.19}$$

where $\beta_i(x) = \mu_{A_i}(x)$ is the degree of membership or belonging of X into the fuzzy set A_i .

The conjunctive form of the antecedent 6.17 can be also formulated as follows,

IF x_1 is $A_{i,1}$ AND ... AND x_p is $A_{i,p}$ THEN $y_i = a_i^\top x + b_i$

with degree of membership or fulfillment

$$\beta_i(x) = \mu_{A_{i,1}}(x_1) \wedge \mu_{A_{i,2}}(x_2) \wedge \dots \wedge \mu_{A_{i,p}}(x_p).$$

By normalizing this membership degree of rule *antecedent* with

$$\phi_i(x) = \frac{\beta_i(x)}{\sum_{j=1}^r \beta_j(x)} \quad (6.20)$$

then we can interpret the *affine* Takagi-Sugeno model as the quasilinear model with a dependent input parameter (Wolkenhauer, 2001):

$$y = \left(\sum_{i=1}^r \phi_i(x) a_i^\top \right) x + \sum_{i=1}^r \phi_i(x) b_i = a^\top(x) + b(x). \quad (6.21)$$

Usually it is difficult to implement multidimensional fuzzy sets, therefore the antecedent 6.17 is commonly interpreted as a combination of equations with a one-dimensional fuzzy set for each variable x .

6.3.3 Model Identification

The basic principle of model identification by product space clustering is to approximate a non linear regression problem by decomposing it to several local linear sub-problems described by IF-THEN rules. A comprehensive discussion can be found in Giles and Draeseke (2001).

The identification and estimation of the fuzzy model used for cases of multi-variate data as follows, suppose

$$y = f(x_1, x_2, \dots, x_p) + \epsilon$$

where ϵ is assumed independent, identical and normally distributed.

If the error has a mean of zero, then the fuzzy function interprets the conditional mean of variable output y . Therefore, the assumption now is the use of linear least square as the basis for analysis.

Step 1: For each x_r and y , separately partition n observations of the sample n into fuzzy clusters c_r by using fuzzy clustering (where $r = 1, \dots, p$).

Step 2: Consider all possible combinations of c fuzzy cluster towards the number of variable input p , where:

$$c = \prod_{r=1}^k c_r \quad (6.22)$$

Step 3: Make a model by using data taken from each fuzzy cluster.

$$y_{ij} = \beta_{i0} + \beta_{i1}x_{1ij} + \beta_{i2}x_{2ij} + \dots + \beta_{ip}x_{pij} + \epsilon_{ij} \quad (6.23)$$

where $j = 1, \dots, n; i = 1, \dots, c$.

Step 4: Predict the conditional mean of x by using:

$$\hat{y}_k = \frac{\sum_{i=1}^c (b_{i0} + b_{i1}x_{1k} + \dots + b_{ip}x_{pk})w_{ik}}{\sum_{i=1}^c w_{ik}}; k = 1, \dots, n. \quad (6.24)$$

where

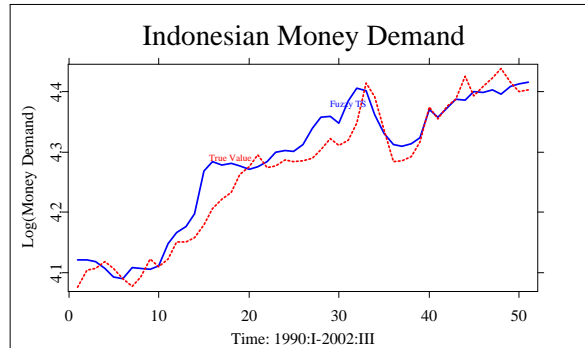
$$w_{ik} = \prod_{r=1}^p \delta_{ij} u_{rjk}; i = 1, \dots, c$$

δ_{ij} is a selector that selects a membership value for each fuzzy of j cluster if the cluster is associated with the i -th (default $\delta_{ij} = 1$).

The fuzzy predictor of the conditional mean y is a weighted average of linear predictor based on the fuzzy partition of explanatory variables, with a membership value varying continuously through the sample observations. The effect of this condition is that the non-linear system can be effectively modelled.

Furthermore, a separately modelling of each fuzzy cluster, including the use of fuzzy logic having the "IF" form, input data are found in this region. "THEN" is likely to be a predictor of the response variables.

The modelling technique based on fuzzy sets can be clustered in a local modelling method, because it uses partitions of a domain process on a number of



Two dimensional plot of Indonesian Money Demand with Fuzzy TS vs Time

fuzzy region. In each region of the input space, a rule is defined to specify models from its output. The rules here are described as a local sub-model of a system. TakagiSugeno is one of the models used here.

6.3.4 Modelling Indonesian Money Demand

In this section we prepare a model for M2 money demand in Indonesia using the approach of fuzzy model identification. This constructing model was based on the three-monthly data taken from 1990:1-2002:III which consisted of three variables, Real Money Demand (M2), Real GNP (GNP), and Long Term Interest Rate (r).

The result of fuzzy clustering with the M2 and GNP variables shows the data formed of three clusters. Of the three clusters, it's difficult to determine their real clusters so that we arrange them as the fourth cluster.

On the other hand, a clustering using M2 and r variables perform of 2 clusters. The intersection of these two clustering stages would result in 4 different clusters as shown in Table 6.2.

The first and the second clusters, that is the period of 1990:1-1994:3 and 1994:4-1998:3, show that GNP has a positive effect of money demand, while r is negative. In the second period, however, the effect of r is not significant. This phenomenon is consistent with the result gained using the standard econometrics method.

Cluster	Observations	Beta 0 (t-value)	Beta 1 - GNP (t-value)	Beta 2 - r (t-value)
<i>I</i>	1-15	-1.7134 (3.402)	0.5479 (5.441)	-0.2047 (-4.149)
<i>II</i>	16-31	0.3237 (0.330)	0.7964 (3.610)	-0.0158 (-0.121)
<i>III</i>	34-39	12.4670 (1.757)	-1.5480 (-1.085)	-0.3177 (-2.377)
<i>IV</i>	40-51	-0.1038 (-0.053)	0.8678 (2.183)	0.1357 (0.901)

Table 6.2: Four clusters of Indonesian money demand data

Similar to the result of standard econometrics method, Fuzzy Takagi-Sugeno approach also performed a coefficient instability during the Asian crisis. Table 6.2 shows that in the third cluster, there is a change in the behavior of the GNP and r in explaining the model of $M2$. During the third period, either the GNP or r has a negative effect toward $M2$. The effect of r is quite significant while the effect of GNP is not. Therefore we may conclude that during the Asian crisis, $M2$ was more influenced by the changing of r .

The last cluster, that is cluster 4, covers the period of 1999:4-2002:3. In this period of time, the situation in Indonesia is slowly under control which was signed by having a new government selected through a democratic general election held in October 1999. In this period, the elasticity of GNP turned back to the level before the time of crisis. Here, the GNP has a positive effect and is significant toward $M2$. On the other hand, the effect of r seemed to be weak and is not significant. This result is also consistent with the finding of the standard econometric one.

6.4 Conclusions

Money demand is generally said to have a functional relationships with the scale variable (representing the economic activity) and the opportunity of holding money balances.

Based on standard econometrics methods, we expect to find equilibrium relationships among these variables. In this paper, we use not only the standard econometrics approach, but also implemented a relatively new approach in the economic analysis called the Fuzzy TakagiSugeno Model.

There was an instability during the Asian crisis, if we observe the situation before the economic crisis (1990 – 1997). During the crisis (1998 – 1999), the effect of GNP and r seemed to be insignificant. After the end of the asian crisis, the effect of GNP tended to be constant and the effect of the interest rate tended to be on the opposite.

Bibliography

- Arize, A. C. (1994). A Re-examination of the Demand for Money in Small Developing Economies, *Applied Economics* **26**: 217-228.
- Arize, A. C. and Shwiff, S. S. (1993). Cointegration, Real Exchange Rate and Modelling the Demand for Broad Money in Japan, *Applied Economics* **25** (6): 717-726.
- Bahmani-Oskooee, M. (2001). How stable is M2 money demand function in Japan?, *Japan and the World Economy* **13**: 455-461.
- Bernanke, B. S. and Gertler, M. (1995). Inside the black box: the credit channel of monetary policy transmission, *Journal of Economic Perspectives* **9**: 27-48.
- Bezdek, J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York.
- Bezdek, J. C. and Pal, S. K. (1992). *Fuzzy Models for Pattern Recognition*, IEEE Press, New York.
- Brand, C. and Cassola, N. (2000). A Money Demand System for Euro Area M3, *ECB Working Paper 39*.
- Coenen, G. and Vega, J. L. (1999). The Demand for M3 in the Euro Area, *ECB Working Paper 6*.
- Deckle, P. and Pradhan, M. (1997). Financial Liberalization and Money Demand in ASEAN Countries: Implications for Monetary Policy. *IMF Working Paper WP/97/36*.
- Drake, L. and Chrystal, K. A. (1994). Company-Sector Money Demand: New Evidence on the Existence of a Stable Long-run Relationship for the UK, *Journal of Money, Credit and Banking* **26**: 479-494.

- Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *Journal of Cybernetics* **3**: 32–57.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and Error Correction: Representation, Estimation and Testing, *Econometrica* **55**: 251–276.
- Ericsson, N. R. (1999). Empirical Modeling of Money Demand, In: Lütkepohl, H. and Wolters, J. (Eds), *Money Demand in Europe*, Physica, Heidelberg, 29-49.
- Giles, D. E. A and Draeseke, R. (2001). Econometric Modelling Using Pattern Recognition via the Fuzzy c-Means Algorithm, in D.E.A. Giles (ed.), *Computer-Aided Econometrics*, Marcel Dekker, New York.
- Goldfeld, S. M. and Sichel, D. E. (1990). The Demand for Money, In: Friedman, B. and Hahn, F. H. (Eds), *Handbook of Monetary Economics*, Elsevier, Amsterdam, 299-356.
- Hafer, R. W. and Jansen, D. W. (1991). The Demand for Money in the United States: Evidence from Cointegration Test, *Journal of Money, Credit and Banking* **23**: 155-168.
- Hafer, R. W. and Kutan, A. M. (1994). Economic Reforms and Long-Run Money Demand in China: Implication for Monetary Policy, *Southern Economic Journal* **60**(4): 936–945.
- Haug, A. A. and Lucas, R. F. (1996). Long-Term Money Demand in Canada: In Search of Stability, *Review of Economic and Statistics* **78**: 345–348.
- Hellendoorn, H. and Driankov, D. (1997). Fuzzy: Model Identification, *Springer Verlag*, Heidelberg.
- Holtemöller, O. (2002). *Vector Autoregressive Analysis and Monetary Policy*. Three Essays, Shaker, Aachen.
- Holtemöller, O. (2003). Aggregation of National Data and Stability of Euro Area Money Demand, In: Institut für Wirtschaftsforschung Halle (Ed), *3rd Workshop in Macroeconometrics*, in print.
- Holtemöller, O. (2004). A Monetary Vector Error Correction Model of the Euro Area and Implications for Monetary Policy, *Empirical Economics*, forthcoming.

- Lim, G. C. (1993). The Demand for the Components of Broad Money: Error Correction and Generalized Asset Adjustment System, *Applied Economics* **25(8)**: 995–1004.
- Lütkepohl, H. and Wolters (Eds)(1999). *Money Demand in Europa*, Physica, Heidelberg.
- McCandless, G. T. and Weber, W. E. (1995). Some Monetary Facts, *Federal Reserve Bank of Minneapolis Quarterly Review* **19**: 2–11.
- McNown, R. and Wallace, M. S. (1992). Cointegration Test of a Long-Run Relation between Money Demand and Effective Exchange Rate, *Journal of International Money and Finance* **11(1)**: 107–114.
- Mehra, Y. P. (1993). The Stability of the M2 Money Demand Function: Evidence from an Error-Correction Model, *Journal of Money, Credit and Banking* **25**: 455–460.
- Miller, S. M (1991). Monetary Dynamics: An Application of Cointegration and Error-Correction Modelling, *Journal of Money, Credit and Banking* **23**: 139–168.
- Mishkin, F. S. (1995). Symposium on the monetary transmission mechanism, *Journal of Economic Perspectives* **9**: 3–10.
- Miyao, R. (1996). Does a Cointegrating M2 Demand Relation Really Exist in Japan?, *Journal of the Japanese and International Economics* **10**: 169–180.
- Moosa, I. A. (1992). The Demand for Money in India: A Cointegration Approach, *The Indian Economic Journal* **40(1)**: 101–115.
- Mucha, H. J. and Sofyan, H. (2000). Cluster Analysis, in Härdle, W., Klinke, S. and Hlavka, Z. *XploRe Application Guide*, (eds.), Springer-Verlag, Heidelberg, Germany
- Nelson, C. R. and Plosser, C. I. (1982). Trends and Random Walks in Macroeconomic Time Series, *Journal of Monetary Economics* **10**: 139–162.
- Orden, D. and Fisher, L. A. (1993). Financial Deregulation and the Dynamics of Money, Prices and Output in New Zealand and Australia, *Journal of Money, Credit and Banking* **25**: 273–292.
- Perron, P. (1989). The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica* **57**: 1361–1401.

-
- Price, S. and Insukindro (1994). The demand for Indonesian narrow money: long-run equilibrium, error correction and forward-looking behaviour, *The Journal of International Trade and Economic Development* **3(2)**: 147–163.
- Ruspini, E. H. (1969). A New Approach to Clustering, *Information Control* **15**: 22–32.
- Sriram, S. S. (1999). Demand for M2 in an Emerging-Market Economy: An Error- Correction Model for Malaysia, *IMF Working paper WP/99/173*.
- Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its application to modelling and control, *IEEE Trans. Systems, Man and Cybernetics* **15(1)**: 116–132.
- Tseng, W. (1994). Economic Reform in China: A New Phase, *IMF Occasional Paper* **114**.
- Walsh, C. E. (1998). *Monetary Theory and Policy*, MIT Press, Cambridge.
- Wolkenhauer, O. (2001). *Data Engineering, Fuzzy Mathematics in system theory and data analysis*, Wiley-Interscience.

7 The exact LR test of the scale in the gamma family

Milan Stehlík

7.1 Introduction

An arbitrary speed asymptotics oversize or undersize the test what may not be acceptable in some fine studies, especially in medicine, finance and physics. In these areas small samples occurs frequently and testing based on asymptotics can be misleading. The high efficiency is typical requirement of testing and exact tests can be asymptotically optimal in the sense of Bahadur, Hodges-Lehman and Pittman. The class of such statistics is apparently narrow, though it contains under certain conditions the likelihood ratio (LR) statistics. Now we are in better situation as statisticians have been in 40s when the one of the main reasons of asymptotics was the fact that the computation of the quantiles of exact tests had the unreal time complexity. Nowadays, the implementation of the special functions to computational environments and symbolic languages became the common phenomenon. The computational and time complexity of procedures of the exact and asymptotical is comparable.

Our setup considers univariate random variables y_1, y_2, \dots, y_N distributed according to the gamma densities

$$f(y_i|\gamma_i) = \begin{cases} \gamma_i^v \frac{y_i^{v-1}}{\Gamma(v)} e^{-\gamma_i y_i}, & \text{for } y_i > 0, \\ 0, & \text{for } y_i \leq 0. \end{cases} \quad (7.1)$$

Here $\gamma_i > 0$, $i = 1, \dots, N$ are unknown scale parameters, which are the parameters of interest and $v > 0$ is known shape parameter.

Such model has been applied in many disciplines, including queueing systems, reliability models, and quality control studies.

We consider the exact LR tests of the scale hypothesis

$$H_0 : \gamma = \gamma_0 \text{ versus } H_1 : \gamma \neq \gamma_0 \quad (7.2)$$

in the scale-homogeneous gamma family (i.e. $\gamma_1 = \dots = \gamma_N := \gamma$). Details can be found in Stehlík (2000), Stehlík (2001) and Stehlík (2002). The general LR test of the hypothesis (2) is implemented to the XploRe in the quantlet `etgammaexacttest`. Because the two special cases of the gamma process, Poisson and Erlang processes are frequently used, they are implemented in self-contained quantlets `etexpexacttest` and `eterlangexacttest`. The advantages of the exact tests of the scale in the gamma family are exact levels of significance (the asymptotical tests are oversized) and exact power. Furthermore, the exact LR tests of the scale in the gamma family are unbiased, uniformly most powerful (UUMP) tests. The theoretical justification of the exact LR testing is its asymptotical optimality. For example, the exact LR test of the scale of the exponential distribution is asymptotically optimal in the sense of the Bahadur exact slopes.

The chapter is organized as follows. In subchapter *Computation the exact tests in the XploRe* we provide the examples of exact testing of the scale parameter. In subchapter *Illustrative examples* we discuss and analyze data set from time processing and field reliability data with missing failure times. To maintain the continuity of the explanation, some theoretical aspects of the exact LR testing are included into the Appendices.

The short discussion on the implementation can be found in Appendix *Implementation to the XploRe*. The discussion on the asymptotical optimality in the case of exponentially distributed observations is given in Appendix *Asymptotical optimality*. Appendix *Information and exact testing in the gamma family* briefly explains the informative aspects of the exact LR testing in the gamma family based on the concept of the "I-divergence distance" between the observation and the canonical parameter γ as presented in Pázman (1993).

Oversizing of the asymptotics provides the Table of oversizing of the asymptotical test. Some useful properties of the Lambert W function, which is crucial function for the implementation of the exact tests is given in Appendix *The Lambert W function*.

7.2 Computation the exact tests in the XploRe

The quantlets `etgammaexacttest`, `etexpexacttest` and `eterlangexacttest` are implemented to the XploRe. The input of the general quantlet `etgammaexacttest` consists of the parameters *sample*, *level*, *nullpar* and *shapepar*. The parameter *sample* inputs the $n \times 1$ data vector x , *level* inputs the size (level of significance) of the exact test, *nullpar* is the input of the value γ_0 of the null hypothesis $H_0 : \gamma = \gamma_0$ and *shapepar* inputs the shape parameter of the gamma distribution. The output consists of the parameters *pvalue*, *testresult*, *chipvalue* and *oversizing*. The parameter *pvalue* outputs the exact p -value of the test, the *test result* outputs the statement about the rejection or acceptance of the null hypothesis. E.g. in the case of rejection this statement has form “ H_0 is rejected on the level alpha“. The parameter *chipvalue* contains the p -value of the χ^2 -Wilks asymptotical test and the *oversizing* contains the oversizing of the χ^2 -asymptotics, i.e. the difference of the *pvalue* and *chipvalue*.

The following session illustrate the application of the quantlet `etgammaexacttest` for exact LR testing of the hypothesis $H_0 : \gamma = 0.4$ versus $H_1 : \gamma \neq 0.4$ at the level of significance $\alpha = 0.1$ for the data set $x = \{0.5458, 2.4016, 1.0647, 2.8082, 0.45669, 0.79987, 0.59494\}$ in the model (1) with the shape parameter equal to 2.3. The libraries `stats` and `nummath` must be loaded before the running of the exact tests.

The result is displayed in Table 7.1.

Table 7.1: Exact LR test $H_0 : \gamma = 0.4$ vs. $H_1 : \gamma \neq 0.4$ with $\alpha = 0.1$

p. value	=	1.0002e-06
test result	=	H_0 is rejected on the level alpha
chip value	=	8.8401e-07
oversizing	=	1.1623e-07

The null hypothesis is rejected on the level $\alpha = 0.1$, the exact p -value is 1.0002e-06 and the asymptotical p -value is 8.8401e-07. The exact p -value is computed from the formula $p = 1 - F_N(-2 \ln \lambda(y))$ where

$$\lambda(y) = \frac{\max_{\gamma \in \Gamma_0} \prod_{i=1}^N f(y_i | \gamma_i)}{\max_{\gamma \in \Gamma_1} \prod_{i=1}^N f(y_i | \gamma_i)} \quad (7.3)$$

is the likelihood ratio of the test of the hypothesis (2), $\Gamma_0 = \{\gamma_0\}$ and $\Gamma_1 = \mathbf{R}^+ \setminus \Gamma_0$. Function F_N is the exact cdf of the Wilks statistics $-2 \ln \lambda$ of the LR test of the hypothesis (2) under the H_0 . The following theorem giving the formula for F_N is the corollary of **Theorem 4** in Stehlík (2001).

Theorem 1 *The exact cdf of the Wilks statistics $-2 \ln \lambda$ of the LR test of the hypothesis (2) of the gamma distribution (1) has under the H_0 the form*

$$F_N(\rho) = \begin{cases} \mathcal{F}_{vN}(-vNW_{-1}(-e^{-1-\frac{\rho}{2vN}})) - \mathcal{F}_{vN}(-vNW_0(-e^{-1-\frac{\rho}{2vN}})), & \rho > 0, \\ 0, & \rho \leq 0. \end{cases}$$

Here W_k , $k = \dots - 1, 0, 1, \dots$ is the k -th branch of the Lambert W function (see Appendix *The Lambert W function*) and \mathcal{F}_N is the cdf of the gamma distribution with the shape parameter $N = 1, 2, \dots$ and scale parameter 1. The Wilks statistics $-2 \ln \lambda(y)$ has under the H_0 asymptotically χ_1^2 -distribution (Wilks, 1967). The *chipvalue* is the p -value obtained from the asymptotics, e.g. $p = 1 - F_{\chi_1^2}(-2 \ln \lambda(y))$ where $\lambda(y)$ is defined by (3) and $F_{\chi_1^2}$ is the cdf of the χ_1^2 -distribution.

7.3 Illustrative examples

7.3.1 Time processing estimation

In this section we give some idea how to analyze the time processing data given in study Gautam (1999). Much more detailed discussion is given in Stehlík (2002). Gautam (1999) studied a typical small scale company which manufactures a single item on demand and according to the specifications of the user. Assume that the users can demand only a single unit of the product. As soon as a demand for a customized item comes in, the company registers the requirements into a database. The processing begins by scanning the database according to a first come first served manner and selecting an item to be processed. Once the processing is complete, the next item's processing begins. Table 7.2 lists the sample processing times in hours.

The following two tasks are of prime concern for the company:

Table 7.2: The sample processing times in hours

1.9174	1.0866	0.9324	1.7576	0.5757	1.3961	1.3726	1.1121	1.3469	1.7477
1.2796	1.5442	1.3416	1.8200	1.0853	1.4793	0.9549	0.8739	1.1211	1.9341
1.6733	1.5880	1.1860	1.3422	1.6634	1.6364	2.7820	1.3701	1.6089	1.5860
1.1534	1.3417	1.6009	1.6476	1.3069	1.4546	1.7063	1.4398	1.1611	1.1278
2.0458	1.3681	0.9886	1.9371	1.6347	1.3754	2.0093	1.2315	1.4088	1.0767
1.0636	1.8241	1.3653	1.3242	1.5804	1.3875	1.2286	1.2866	1.7617	1.4828
1.6378	0.8287	1.1559	1.1054	1.4677	1.1568	1.6271	1.3273	1.5050	1.4858
1.1695	2.3610	1.3828	2.0455	1.1538	1.0897	1.7877	1.2850	1.3557	2.1030
1.0352	1.6838	1.6230	0.8023	1.6138	1.7363	0.8154	1.1386	2.1088	1.0676
1.5356	1.7195	1.9940	1.7607	1.2447	1.4794	1.1470	1.4041	1.1489	1.7280

- Given the number of orders waiting to be processed, by what time should the company promise to deliver the product so that company meets the deadline with the probability p ?
- Consider that a job was being processed for the past u hours when the system breaks down. When the breakdown is fixed, the company would like to estimate how much longer the processing will take.

One of the probability distributions representing time random variable with relatively small variation is the Erlang distribution. The family of density functions

$$f(y|\gamma) = \frac{\gamma(\gamma y)^{k-1}}{(k-1)!} e^{-\gamma y}, \quad y > 0, \quad k = 2, 3, \dots$$

is referred to as the family of Erlang (k, γ) distributions. The length of time interval required in order to collect k arrivals from a Poisson process is distributed according to the Erlang distribution with the shape parameter k . The family of Erlang distributions with the known shape parameter $v = k > 1$ and an unknown scale parameter γ belongs to the model (1). The quantlets `cdf Erlang` and `pdf Erlang` implemented to the XploRe computes the values of the cdf and pdf of the Erlang distribution with called parameters.

To fit the processing time data as Erlang (k, γ) (Gautam, 1999) estimates the parameters from the data and test if the estimated distribution fits the data well. He uses the maximum likelihood estimators (MLEs) $\hat{k} = \lfloor \frac{m^2}{s^2} \rfloor$ or $\lfloor \frac{m^2}{s^2} \rfloor + 1$ and $\hat{\gamma} = \frac{m}{s^2}$ and where m and s^2 denotes the sample mean and sample variance

and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x . The estimation from the data in Table 1 gives MLEs $\hat{\gamma} = 11.1992$ and $\hat{k} = 16$ or 17. Gautam (1999) use the Kolmogorov-Smirnov (KS) test to test the goodness-of-fit of the estimated distribution. The KS test states that the null hypothesis cannot be rejected at the level 0.05 in the both Erlang(16, 11.1992) and Erlang(17, 11.1992) cases. Consider that there are m orders to be processed when a customer request comes in. Denote X_1, \dots, X_{m+1} the processing times all according to an Erlang(k, γ). Denote Z the time for the $m + 1$ jobs to be completed. We have $Z = X_1 + X_2 + \dots + X_{m+1}$ and Z has Erlang($k(m + 1), \gamma$) distribution.

Promising a due date

Our problem is, what time should the company promise to deliver the product so that it meets the deadline with the given probability p (let us say 0.99)? We have to determine time T such that $P\{Z \leq T\} = 0.99$. We can solve T from the equation

$$1 - e^{-\gamma T} \sum_{i=0}^{k(m+1)-1} \frac{(\gamma T)^i}{i!} = 0.99. \quad (7.4)$$

In the case of Erlang(16, 11.1992) distributed processing times (the situation corresponding to the dataset in the Table 1), $m = 4$ and $p = 0.99$ we obtain $T = 9.1315$ hours.

Assume that the processing time in the company has Erlang distribution with known shape parameter $k = 16$ and unknown scale parameter γ . The customer wants to know the time to deliver and in the case of delay larger than one hour the company will be penalized. To determine the uncertainty of the penalization we want to test $H_0 : T = 9.1315$ versus $H_1 : T \neq 9.1315$ at the level of significance $\alpha = 0.05$ having large power of this test at the point of penalization 10.0938 of the alternative.

For γ and T the equality (4) holds, we have the correspondence $\gamma = \gamma(T)$ and $11.1992 = \gamma(9.1315)$ and $10.0938 = \gamma(10.1315)$ holds. Therefore the corresponding scale testing problem is

$$H_0 : \gamma = 11.1992 \text{ versus } H_1 : \gamma \neq 11.1992. \quad (7.5)$$

The relevant idea is to test the hypothesis (5) exactly. The asymptotical Wilks test is oversized and the power is smaller than the exact one. The significance of this effect can be clearly seen from the Table 3 in Appendix *Oversizing of the asymptotics*. The value of the Wilks statistics of the LR test of the hypothesis (6) is $-2 \ln \lambda = 0.14981$ and the corresponding p -value is 0.69873. Therefore the null hypothesis cannot be rejected at the 0.05 significance level. Much more detailed discussion can be found in Stehlík (2002), where the exact power function is derived (in **Theorem 1**):

The exact power $p(\gamma, \alpha)$ of the LR test based on the Wilks statistics of the hypothesis (2) of the Erlang(k, γ) distribution on the level α at the point γ of the alternative has the form

$$p(\gamma, \alpha) = 1 - \mathcal{F}_{kN} \left(-kN \frac{\gamma}{\gamma_0} W_{-1} \left(-e^{-1 - \frac{c_{\alpha, N}}{2kN}} \right) \right) + \mathcal{F}_{kN} \left(-kN \frac{\gamma}{\gamma_0} W_0 \left(-e^{-1 - \frac{c_{\alpha, N}}{2kN}} \right) \right), \quad (7.6)$$

where $c_{\alpha, N}$ denotes the critical value of the exact test of the hypothesis (2) on the level α .

We obtain the exact power $p(10.0938, 0.05) = 0.9855$. Since the power of our test is high at the point corresponding to the penalization time, the probability of penalization is small. The power function $p(\gamma, 0.05)$ of the LR test of the hypothesis (5) is for $\gamma \in (7, 15)$ displayed in the Figure 1.

Time to complete after breakdown

Consider that a job was being processed for the past u hours when the system breaks down. When the breakdown is fixed, the company would like to estimate how much longer the processing will take. Denote $p(t)$ the probability $P\{X \leq u+t | X > u\}$ (X denotes the processing time) that once the breakdown is fixed, the processing will continue for less than t hours given that the breakdown occurred u hours after the processing started. It is easy to see that

$$p(t) = 1 - e^{-\gamma t} \frac{\sum_{i=0}^{k-1} \frac{(\gamma(u+t))^i}{i!}}{\sum_{i=0}^{k-1} \frac{(\gamma u)^i}{i!}}.$$

Consider the processing times given from the Table 7.2. Then the processing time has Erlang (16, 11.1992) distribution. If a system breakdown occurs 1.5 hour after processing starts, then the probability $p(0.5)$ that the processing will

Table 7.3: Airplane indicator light reliability data

Failures	Cumulative operating time(hours)
2	$T_{21} = 51000$
9	$T_{91} = 194900$
8	$T_{81} = 45300$
8	$T_{82} = 112400$
6	$T_{61} = 104000$
5	$T_{51} = 44800$

be complete within 30 minutes after the breakdown is fixed equals to 0.83079. The application of the exact LR test of the scale would be similar like in the part *Promising a due date*.

7.3.2 Estimation with missing time-to-failure information

For the assessment of component reliability, field data has many distinct advantages (Coit and Jint, 2000). For all of the advantages of the field data, there are also disadvantages, including incomplete or inaccurate data reporting and others. Several of these disadvantages are described in more detail by Coit, Dey and Jint (1986). The disadvantage to be addressed in Coit and Jint (2000) is the fact that the individual times-to-failure are often missing. The data is often only available in the form of r collective failures observed T_{rj} cumulative hours with no further delineation or detail available (Anon, 1991) and (Anon, 1997). Quantities r and T_{rj} are known but the individual failure times are not. Analysts may have many of these merged data records available for the same component. Table 7.4 presents a data set of this type. Here T_{rj} is the j th cumulative operating time with r failures, i.e.

$$T_{rj} = X_1 + \dots + X_r,$$

where X_i is the i th time-to-failure.

There has been other research concerned with the use of data with missing attributes. Dey (1982) has developed a simulation model to observe the behavior of grouped data and test an exponential distribution assumption. Coit and Dey (1999) have developed and demonstrated a hypothesis test to evaluate an

exponential distribution assumption when there is missing time-to-failure data. The grouped exponential data was modelled using a k -Erlang distribution (Coit and Jint, 2000). The hypothesis test was demonstrated to successfully reject the exponential distribution when it was not appropriate even without detailed knowledge of component time-to-failure. Coit and Jint (2000) made following assumptions:

- Component times-to-failure are *iid*
- Component time-to-failure are gamma distributed
- Repair times are insignificant compared to operating time
- System repair does not degrade or otherwise affect the reliability of the unfailed components.

Time to failure distributions

For standard data sets, there are well-known techniques to estimate parameters for many distributions and types of censoring, and to objectively evaluate the applicability of these distributions. However, the analysis of reliability data with missing time-to-failure is a nonstandard problem and it is not addressed in these references.

Without individual time-to-failure data, it is impossible to fit the data to most popular distributions (e.g. gamma, Weibull, lognormal) using standard techniques such as MLE. However, the MLE for the exponential distribution with Type I or II censoring only depends on the number of failures and the cumulative times.

This is a product of the memoryless property associated with the exponential distribution. The limitations of the available field data and the simplicity of the exponential MLE have been used to rationalize the exponential distribution in applications where it would seemingly be a poor choice. the constant hazard function associated with the exponential distribution is not intuitively appropriate for most failure mechanisms which can be attributed to the accumulation of stress, such as fracture, fatigue, corrosion and wear mechanisms.

Incorrect assumptions of the underlying distribution can have dire consequences. For many fledging companies, major decisions are made with limited data being used as rationale. When an incorrect distribution is assumed, particularly for reasons of convenience, it is particularly dangerous.

The gamma distribution is flexible distribution that can model many particular component failure mechanisms. Methods to determine MLE estimates are available (Lawless, 1982) for Type I, Type II and chronologically grouped failure data. However, none of the available MLE estimators pertain to the data with missing failure times.

Gamma-distribution maximum-likelihood estimates

- **Distribution of cumulative time-to-failure**

Coit and Jint (2000) determine gamma distribution MLEs with missing failure times. The likelihood function is based on the distribution parameters (γ, v) and observed r and T_{rj} values rather than times-to-failure and censor times which are unknown. When time-to-failure X_i is distributed according to the gamma distribution $Gamma(\gamma, v)$ then $T_{rj} = \sum_{i=1}^r X_i$ is distributed in accordance with a related gamma distribution $Gamma(\gamma, rv)$. T_{rj} is a random variable depending on specified r . If r is also a random variable, then consider T as the associated cumulative time-to-failure for random r . The density function for T can be expressed as a sum of conditional terms. We have

$$f_{T_r}(t) = f_T(t|r) = \frac{\gamma^{rv} t^{rv-1}}{\Gamma(rv)} e^{-\gamma t}$$

and

$$f_T(t) = \sum_{r=1}^{\infty} f_T(t|r)P(R = r),$$

where $P(R = r)$ is the probability that a randomly selected data record includes r failures.

Consider a specific observed data set with N data records. For each of the N data records, r and T_{rj} are known. The set of cumulative operating times $\{T_{rj}\}$ form a non-homogeneous population that can be modelled using a mixture model. The T_{rj} variables are associated with different gamma distributions depending on r with the same scale parameter. Mixture models are generally used when there is a nonhomogeneous population composed of distinct subpopulations. The most common situation is when the proportion of each subpopulation is known or can be estimated, but it is unknown which members belong to which subpopulations. The probability density function for a non-homogeneous population can be expressed as weighted sum of the respective probability density functions. The weights represent the probability that

a randomly selected member of the population is from a particular subpopulation. For our problem, the subpopulations are characterized by the number of failures within the merged data set. Denote n_r the number of data records with exactly r failures. The weights are the probabilities that a randomly selected T_{rj} from the N data records has r failures. This probability is $\frac{n_r}{N}$. Denote $\mathbf{n} = (n_1, n_2, \dots, n_m)$ where $m = \max\{r | n_r > 0\}$ is the maximum number of failures for any data record within a data set and $N = \sum_{r=1}^m n_r$. Thus we have

$$f_T(t|\mathbf{n}) = \frac{1}{N} \sum_{r=1}^m \frac{n_r \gamma^{rv} t^{rv-1}}{\Gamma(rv)} e^{-\gamma t}. \quad (7.7)$$

• **Maximum likelihood estimate**

A likelihood function for γ and v is based on the observed data and relation (7). The likelihood function $l(v, \gamma)$ is expressed as a product of $f_T(t|\mathbf{n})$ for the N data records

$$l(v, \gamma) = \prod_{r=1}^m \prod_{j=1}^{n_r} \frac{1}{N} \sum_{k=1}^m \frac{n_k \gamma^{kv} T_{rj}^{kv-1}}{\Gamma(kv)} e^{-\gamma T_{rj}}.$$

Estimates of v and γ can be obtain using a Newton search to minimize $l(v, \gamma)$. An alternative and more preferred likelihood function can be developed by exploiting the observation that the subpopulations are clearly identified within the overall population. The alternative likelihood function can be expressed as the product of the conditional density functions for T_{rj} and $f_T(t|r)$ as follows

$$l_2(v, \gamma) = \prod_{r=1}^m \prod_{j=1}^{n_r} f_T(v, \gamma | T_{rj})$$

and

$$\ln l_2(v, \gamma) = -\gamma M \bar{t} + Mv \ln \gamma + \sum_{r=1}^m \sum_{j=1}^{n_r} (vr - 1) \ln T_{rj} - \sum_{r=1}^m n_r \ln \Gamma(rv), \quad (7.8)$$

where $M = \sum_{r=1}^m r n_r$ denotes total number of failures associated with all N data records and $\bar{t} = \frac{1}{M} \sum_{r=1}^m \sum_{j=1}^{n_r} T_{rj}$ is average time-to-failure.

Taking partial derivatives of the log-likelihood function $\ln l_2$ gives the equations

$$\hat{\gamma} = \frac{\hat{v}}{\hat{t}} \quad (7.9)$$

and

$$M \ln\left(\frac{v}{\hat{t}}\right) + \sum_{r=1}^m \sum_{j=1}^{n_r} r \ln T_{rj} - \sum_{r=1}^m r n_r \psi(rv) = 0, \quad (7.10)$$

where ψ is Euler's digamma function. The digamma function is in Coit and Jint (2000) approximated by the equation

$$\psi(x) = \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \dots \quad (7.11)$$

Now \hat{v} can be solved from the equations (10) and (11) using a Newton-Raphson or bisection search, \hat{v} is then substituted into equation (9) to determine $\hat{\gamma}$. The resulting gamma estimates should be tested for goodness-of-fit.

- **Testing the scale of the failure process**

Suppose, that the shape parameter is v and consider the test of the hypothesis (2):

$$H_0 : \gamma = \gamma_0 \text{ versus } H_1 : \gamma \neq \gamma_0$$

about the scale of the failure process. Such testing problem is usual in analysis of failure processes. Suppose that we have the only one observation T_{r1} which is typical in the case of unsystematic rebooting of some device. Then T_{r1} has $Gamma(\gamma, rv)$ distribution and the optimal test is the exact LR test of the scale in the gamma family. We use the quantlet `etgammaexacttest` in the form `etgammaexacttest(x, alpha, gamma_0, rv)` where $\mathbf{x} = \{T_{r1}\}$ is data vector. The power function can be evaluated with the help of the following theorem:

Theorem 2 *The exact power $p(\gamma, \alpha)$ of the LR test based on the Wilks statistics of the hypothesis (2) of the $Gamma(v, \gamma)$ distribution on the level α at the point γ of the alternative has the form*

$$p(\gamma, \alpha) = 1 - \mathcal{F}_{vN} \left\{ -vN \frac{\gamma}{\gamma_0} W_{-1} \left(-e^{-1 - \frac{c_{\alpha, N}}{2vN}} \right) \right\} + \mathcal{F}_{vN} \left\{ -vN \frac{\gamma}{\gamma_0} W_0 \left(-e^{-1 - \frac{c_{\alpha, N}}{2vN}} \right) \right\},$$

where $c_{\alpha, N}$ denotes the critical value of the exact test of the hypothesis (2) on the level α .

Proof. The critical region based on the Wilks statistics of the LR test of the hypothesis (2) on the level of significance α has the form

$$W_c = \{y \in Y : -2 \ln \lambda(y) > c\} \text{ such that } P\{W_c | \gamma = \gamma_0\} = \alpha, \quad (7.12)$$

where Y denotes the sample space. The power $p(\gamma_1, \alpha)$ of the test of the hypothesis (2) at the point γ_1 of the alternative is equal to $P\{W_c | \gamma = \gamma_1\}$, where W_c is defined by (12).

Applying **Theorem 4** in (Stehlík, 2001) we obtain in the case of the Gamma(v, γ) distributed observations the equality

$$1 - P\{W_c | \gamma = \gamma_1\} = \mathcal{F}_{vN} \left\{ -vN \frac{\gamma_1}{\gamma_0} W_{-1} \left(-e^{-1 - \frac{c_{\alpha, N}}{2vN}} \right) \right\} - \mathcal{F}_{vN} \left\{ -vN \frac{\gamma_1}{\gamma_0} W_0 \left(-e^{-1 - \frac{c_{\alpha, N}}{2vN}} \right) \right\}.$$

This completes the proof. \square

The testing of the scale of the failure process based on the one observation is nice example of application of the testing based on "one measurement" which accumulate information from more failures. In such situations the exact test based on the one measurement can be useful.

The other interesting case of scale testing is sample of the form $T_{r1}, T_{r2}, \dots, T_{rN}$ where r is fixed number of observed failures. Then T_{ri} , $i = 1, \dots, N$ are *iid* and Gamma(γ, rv) distributed.

The optimal test of the hypothesis (2) is the exact LR test of the scale in the gamma family, we use the quantlet `etgammaexacttest` again.

The exact test for the general form $\{T_{r,j}\}$ of sample is much more complicated and can be found in Stehlík (2002a).

7.4 Implementation to the XploRe

The cdfs of the Wilks statistics of the LR tests of the scale are written in terms of the incomplete Gamma function and two branches W_0 and W_{-1} of the LW function. The branches of the LW function are implemented to the standard

mathematical computational environments, e.g. Maple, Matlab, Mathematica and Mathcad. The reason of this implementation is a large diversity of problems to their the LW gives solution. If the LW is not implemented in some statistical software (which is the case of the XploRe) we can proceed as follows. Let us consider the selectors $\min(S)$ and $\max(S)$, where S is the set of solutions of the equation $x - \ln(x) = c$, $c > 1$. Then we have the following relations in the terms of the Lambert W function (Stehlík, 2001) for proof :

$$\min(S) = -W_0(-e^{-c})$$

and

$$\max(S) = -W_{-1}(-e^{-c}).$$

Practically, we use the numerical solver of the equation $x - \ln(x) = c$. The problem is that the derivative is unbounded at the zero neighborhood. Thus the speed methods as Newton-Raphson or modified Newton-Raphson based on derivative are not applicable. We use the quantlet `nmbrentroot.xpl` (Čížková, 2003). The `nmbrentroot.xpl` is the XploRe implementation of the Brent's improvement (Press, 1992) of the van Wijngaarden-Dekker method. Brent's method joint the advantages of the root bracketing, bisection method and inverse quadratic interpolation and this is the way to combine superlinear convergence with the sureness of bisection.

7.5 Asymptotical optimality

The one of the much more interesting theoretical properties of the exact LR test of the scale of the exponential distribution is its asymptotical optimality (AO) in the sense of Bahadur, which can be described as follows.

Consider a testing problem $H_0 : \vartheta \in \Theta_0$ vs $H_1 : \vartheta \in \Theta_1 \setminus \Theta_0$, where $\Theta_0 \subset \Theta_1 \subset \Theta$. Further consider sequence $T = \{T_N\}$ of test statistics based on measurements y_1, \dots, y_N which are iid according to an unknown member of an family $\{P_\vartheta : \vartheta \in \Theta\}$. We assume that large values of test statistics give evidence against H_0 . For ϑ and t denote $F_N(t, \vartheta) := P_\vartheta\{s : T_N(s) < t\}$ and $G_N(t) := \inf\{F_N(t, \vartheta) : \vartheta \in \Theta_0\}$. The quantity $L_n(s) = 1 - G_n(T_n(s))$ is called the attained level or the p -value. Suppose that for every $\vartheta \in \Theta_1$ the equality

$$\lim_{n \rightarrow \infty} \frac{-2 \ln L_n}{n} = c_T(\vartheta)$$

holds a.e. P_ϑ .

Then the nonrandom function c_T defined on Θ_1 is called the Bahadur exact slope of the sequence $T = \{T_n\}$. According to the theorem of Raghavachari (1970) and Bahadur (1971), the inequality

$$c_T(\vartheta) \leq 2K(\vartheta, \Theta_0) \quad (7.13)$$

holds for each $\vartheta \in \Theta_1$. Here $K(\vartheta, \Theta_0) := \inf\{K(\vartheta, \vartheta_0) : \vartheta_0 \in \Theta_0\}$ and $K(\vartheta, \vartheta_0)$ denotes the Kullback-Leibler information number. If (13) holds with the equality sign for all $\vartheta \in \Theta_1$, then the sequence T is said to be asymptotically optimal in the Bahadur sense. The maximization of $c_T(\vartheta)$ is a nice statistical property, because the greater the exact slope is, the more one can be convinced that the rejected null hypothesis is indeed false. The class of such statistics is apparently narrow, though it contains under certain conditions the LR statistics (Bahadur, 1967) and (Rublík, 1989a).

Rublík (1989a) proved AO of the LR statistic under regularity condition which is shown in (Rublík, 1989b) to be fulfilled by regular normal, exponential and Laplace distribution under additional assumption that Θ_0 is a closed set and Θ_1 is either closed or open in metric space Θ . For more extensive discussion on asymptotical optimality see also monograph of Nikitin (1995).

7.6 Information and exact testing in the gamma family

Model (1) is a regular exponential family (Barndorff-Nielsen, 1978), the sufficient statistics for the canonical parameter $\gamma \in \Gamma$ has the form $t(y) = -y$ and $\Gamma = \{(\gamma_1, \dots, \gamma_N), \gamma_i > 0; i = 1, \dots, N\}$. The "covering" property

$$\{t(y) : y \in Y\} \subseteq \{E_\gamma[t(y)] : \gamma \in \Gamma\}$$

(Pázman, 1993) together with the relation

$$E_\gamma[t(y)] = \frac{\partial \kappa(\gamma)}{\partial \gamma}$$

where $\kappa(\gamma) = N \ln(\Gamma(v)) - v \sum_{i=1}^N \ln(\gamma_i)$ enables us to associate to each value of $t(y)$ a value $\hat{\gamma}_y \in \Gamma$ which satisfies

$$\frac{\partial \kappa(\gamma)}{\partial \gamma} \Big|_{\gamma = \hat{\gamma}_y} = t(y). \quad (7.14)$$

The equation (14) follows that $\hat{\gamma}_y$ is the MLE of the canonical parameter γ in the family (1). By the use of (14) we can define the I -divergence of the observed vector y in the sense of Pázman (1993):

$$I_N(y, \gamma) := I(\hat{\gamma}_y, \gamma).$$

Here $I(\gamma^*, \gamma)$ is the Kullback-Leibler divergence between the parameters γ^* and γ . The I -divergence has nice geometrical properties, let us mention only the Pythagorean relation

$$I(\bar{\gamma}, \gamma) = I(\bar{\gamma}, \gamma^*) + I(\gamma^*, \gamma)$$

for every $\gamma, \bar{\gamma}, \gamma^* \in \text{int}(\Gamma)$ such that $(E_{\bar{\gamma}}(t) - E_{\gamma^*}(t))^T(\gamma^* - \gamma) = 0$. The Pythagorean relation can be used for construction of the MLE density in regular exponential family, see Pázman (1996) for details.

The I -divergence have also nice statistical consequences. Let us consider the LR λ_1 of the test of the hypothesis (2) and the LR λ_2 of the test of the homogeneity hypothesis $H_0 : \gamma_1 = \dots = \gamma_N$ in the family (1).

Then we have the following interesting relation for every vector of canonical parameters $(\gamma_0, \dots, \gamma_0) \in \Gamma^N$:

$$I_N(y, (\gamma_0, \dots, \gamma_0)) = -\ln \lambda_1 + (-\ln \lambda_2 | \gamma_1 = \dots = \gamma_N). \quad (7.15)$$

Here the variables $-\ln \lambda_1$ and $-\ln \lambda_2$ under the condition $H_0 : \gamma_1 = \dots = \gamma_N$ are independent. The relation (15) is consequence of the **Theorem 4** in Stehlík (2001).

7.7 The Lambert W function

The Lambert W function is defined to be the multivalued inverse of the complex function $f(y) = ye^y$. As the equation $ye^y = z$ has an infinite number of solutions for each (non-zero) value of $z \in \mathbf{C}$,

the Lambert W has an infinite number of branches. Exactly one of these branches is analytic at 0. Usually this branch is referred to as the principal branch of the Lambert W and is denoted by W or W_0 .

The other branches all have a branch point at 0. These branches are denoted by W_k where $k \in \mathbf{Z} \setminus \{0\}$. The principal branch and the pair of branches W_{-1} and W_1 share an order 2 branch point at $z = -e^{-1}$.

The branch cut dividing these branches is the subset of the real line from $-\infty$ to $-e^{-1}$ and the values of the branches of the Lambert W on this branch cut are assigned using the rule of counter-clockwise continuity around the branch point. This means that W is real-valued for $z \in (-e^{-1}, \infty)$ and the branch W_{-1} is real-valued on the interval $(-e^{-1}, 0)$. For all the branches other than the principal branch the branch cut dividing them is the negative real axis. The branches are numbered up and down from the real axis. A detailed discussion of the branches of the Lambert W can be found in Corless, *et.al.* (1996).

Since the Lambert W function has many applications in pure and applied mathematics, the branches of the LW function are implemented to many mathematical computational softwares, e.g. the Maple, Matlab, Mathematica and Mathcad. For more information about the implementation and some computational aspects (Corless, *et.al.*, 1993).

7.8 Oversizing of the asymptotics

The Wilks statistics $-2 \ln \lambda$ of the LR test of the hypothesis (2) is under H_0 asymptotically χ_1^2 -distributed (Wilks, 1967) and the test based on the asymptotics is oversized. The following table gives the oversizing of the asymptotical test for small samples when the observations are distributed exponentially. Here α is the size of the test given from the Wilks asymptotics while $\alpha_{e,N}$ is the exact size of the same test. We calculate from the formula $\alpha_{e,N} = 1 - F_N(\chi_{\alpha,1}^2)$. Here $\chi_{\alpha,1}^2$ denotes $(1 - \alpha)$ -quantile of the asymptotical χ_1^2 -distribution and F_N is the exact cdf of the Wilks statistics $-2 \ln \lambda$ of the LR test of the hypothesis (2) under the H_0 . The formula for F_N is derived in Stehlík (2001) and has the form:

$$F_N(\tau) = \begin{cases} \mathcal{F}_N(-NW_{-1}(-e^{-1-\frac{\tau}{2N}})) - \mathcal{F}_N(-NW_0(-e^{-1-\frac{\tau}{2N}})), & \tau > 0, \\ 0, & \tau \leq 0. \end{cases}$$

Table 7.4: The exact sizes $\alpha_{e,N}$

$\alpha \backslash N$	1	2	3	4	5
0.00001	0.229211e-4	0.178154e-4	0.155230e-4	0.142351e-4	0.134204e-4
0.00002	0.445707e-4	0.347445e-4	0.303720e-4	0.279358e-4	0.264021e-4
0.00005	0.1070863e-3	0.838629e-4	0.736765e-4	0.680620e-4	0.645496e-4
0.0001	0.2073771e-3	0.1630849e-3	0.1439040e-3	0.1334225e-3	0.1268967e-3
0.0002	0.4007327e-3	0.3167122e-3	0.2808525e-3	0.2614285e-3	0.2493896e-3
0.0005	0.9536900e-3	0.7598856e-3	0.6789730e-3	0.6356518e-3	0.6089504e-3
0.001	0.18315866e-2	0.14706397e-2	0.13227325e-2	0.12442198e-2	0.11960162e-2
0.002	0.35061415e-2	0.28417427e-2	0.25749120e-2	0.24344361e-2	0.23484946e-2
0.005	0.82247349e-2	0.67718253e-2	0.62044830e-2	0.59087828e-2	0.57286067e-2
0.01	0.015599286	0.013037809	0.012058871	0.011552053	0.011244013
0.02	0.029448482	0.025065314	0.023424550	0.022579936	0.022067611
0.05	0.067701923	0.059361294	0.056314364	0.054754992	0.053810812
$\alpha \backslash N$	6	7	8	9	10
0.00001	0.128631e-4	0.124589e-4	0.121533e-4	0.119145e-4	0.117233e-4
0.00002	0.253549e-4	0.245975e-4	0.240255e-4	0.235786e-4	0.232200e-4
0.00005	0.621610e-4	0.604363e-4	0.591357e-4	0.581205e-4	0.573069e-4
0.0001	0.1224689e-3	0.1224689e-3	0.1168726e-3	0.1149965e-3	0.1134937e-3
0.0002	0.2412404e-3	0.2353750e-3	0.2309574e-3	0.2275143e-3	0.2247565e-3
0.0005	0.5909296e-3	0.5779784e-3	0.5682339e-3	0.5606426e-3	0.5545638e-3
0.001	0.11635466e-2	0.1140235e-2	0.11227052e-2	0.11090527e-2	0.10981231e-2
0.002	0.22907016e-2	0.22492459e-2	0.22180864e-2	0.21938248e-2	0.21744048e-2
0.005	0.56076661e-2	0.55209911e-2	0.54558724e-2	0.54051793e-2	0.53646073e-2
0.01	0.0110374652	0.0108895082	0.0107783725	0.0106918635	0.010622626
0.02	0.0217243535	0.0214785418	0.0212939228	0.021150213	0.0210351898
0.05	0.0531785833	0.0527258996	0.0523858952	0.0521212038	0.0519093212
$\alpha \backslash N$	11	20	30	40	50
0.00001	0.115663e-4	0.108587e-4	0.105711e-4	0.104276e-4	0.103418e-4
0.00002	0.229270e-4	0.216047e-4	0.210675e-4	0.207998e-4	0.206392e-4
0.00005	0.566409e-4	0.536422e-4	0.524229e-4	0.518152e-4	0.514510e-4
0.0001	0.1122631e-3	0.1067253e-3	0.1044757e-3	0.1033529e-3	0.1026806e-3
0.0002	0.2224988e-3	0.2123424e-3	0.2082143e-3	0.2061548e-3	0.2049209e-3
0.0005	0.5495886e-3	0.5272108e-3	0.5181143e-3	0.5135745e-3	0.5108540e-3
0.001	0.10891785e-2	0.10489474e-2	0.10325913e-2	0.10244260e-2	0.10195322e-2
0.002	0.21585122e-2	0.20870295e-2	0.20579594e-2	0.20434441e-2	0.20347421e-2
0.005	0.53314054e-2	0.51820403e-2	0.51212668e-2	0.50909105e-2	0.50727079e-2
0.01	0.0105659642	0.010311006	0.010207222	0.010155366	0.010124268
0.02	0.0209410536	0.00517344	0.020344780	0.020258532	0.020206799
0.05	0.0517358893	0.050954881	0.050636560	0.050477398	0.050381907

Bibliography

- Anonymous (1991). *Non-electronic Parts Reliability Data*, RAC Report NPRD, Reliability Analysis Center, Rome, NY.
- Anonymous (1997). *Electronic Parts Reliability Data*, RAC Report EPRD, Reliability Analysis Center, Rome, NY.
- Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families In Statistical Theory*, John Wiley & Sons, Ltd., New York, pp. 111-115.
- Bahadur, R. R. (1967). *Rates of convergence of estimates and test statistics*, Ann. Mathem. Statist. **38**, pp. 303-324.
- Bahadur, R. R. (1971). *Some Limit Theorems in Statistics*, Philadelphia, SIAM.
- Čížková, L. (2003). Numerical Optimization Methods in Econometrics, In: Rodriguez-Poo, J. (eds.), *ComputerAided Introduction to Econometrics Guide*, Springer-Verlag.
- Coit, D. W. and Jin, T. (2000). *Gamma distribution parameter estimation for field reliability data with missing failure times*, IIE Transactions **32**, 1161-1166.
- Coit, D. W., Dey, K. A., and Turkowski, W. (1986). *Practical reliability data and analysis*, Reliability Engineering, **14**, 1-18.
- Coit, D. W. and Dey, K. A. (1999). *Analysis of grouped data from field-failure reporting systems*, Reliability Engineering and System Safety, **65**, 95-101.
- Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1996). *On the Lambert W function*, Advances in Computational mathematics **5**, 329-359.

- Corless, R. M., Gonnet, G. H., Hare, D. E. G., Jeffrey, D. J., and Knuth, D. E. (1993). *Lambert's W Function in Maple*, Maple Technical Newsletter **9**, pp. 12-22.
- Dey, K. A. (1982). *Statistical analysis of noisy and incomplete failure data*, in Proceedings Annual Reliability and Maintainability Symposium(RAMS), IEEE, Piscataway, NJ. pp. 93- 97.
- Gaudoin, O. and Soler, J. L. (1992). *Statistical Analysis of the Geometric De-Eutrophication Software Reliability Model*, IEEE Transactions on Reliability, **41**.
- Gautam, N. (1999). *Erlang Distribution: Description and Case Study*, Industrial Engineering Applications and Practice: Users Encyclopedia.
- Kleinrock, L. (1975). *Queueing Systems*, vol. 1, John Wiley & Sons, Toronto, pp. 71-72, 119-134.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, New York.
- Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, New York, pp, 1-36
- Pázman, A. (1995). *Nonlinear statistical Models*, Kluwer Acad. Publ., Dordrecht, ch. 9.1, 9.2
- Pázman, A. (1995). *The Density of the Parameter Estimators when the Observations are Distributed Exponentially*, *Metrika*, **44**, Physica-Verlag, Heidelberg, pp. 9-26.
- Press, W. H., Teukolsky, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C- Second Edition*, Cambridge University Press, Cambridge, pp. 359-362.
- Raghavachari, M. (1992). *On a theorem of Bahadur on the rate of convergence of test statistics*, *Ann. Mathem. Statist.* **41**, pp. 1695-1699.
- Rublík, F. (1989). *On optimality of the LR tests in the sense of exact slopes*, Part 1, General case. *Kybernetika* **25**, pp. 13-25.
- Rublík, F. (1989). *On optimality of the LR tests in the sense of exact slopes*, Part 2, Application to Individual Distributions. *Kybernetika* **25**, pp. 117-135.

-
- Stehlík, M. (2000). *On some exact densities in an exponential family*, in: J. Antoch and G. Dohnal (Eds.) ROBUST, Praha, pp. 298-307.
- Stehlík, M. (2001). *Distributions of exact tests in the exponential family*, Accepted to Metrika.
- Stehlík, M. (2002). *The exact LR test of the scale in the gamma family*, Accepted to Tatra Mountains Math. Publ.
- Stehlík, M. (2002). *The exact test of the scale for field reliability data with missing failure times*, Will appear.
- Wilks, S. S. (1967). *Mathematical Statistics*, Nauka, Moscow, pp. 410-419.

8 Pricing of catastrophe (CAT) bonds

Krzysztof Burnecki, Grzegorz Kukla, David Taylor

8.1 Introduction

9 Extreme value theory

Krzysztof Jajuga, Daniel Papla

9.1 Introduction

10 Applying Heston's stochastic volatility model to FX options markets

Uwe Wystup, Rafal Weron

10.1 Introduction

11 Mortgage backed securities: how far from optimality

Nicolas Gausse, Julien Tamine

11.1 Introduction

12 Correlated asset risk and option pricing

Wolfgang Härdle, Matthias Fengler, Marc Tisserand

12.1 Introduction

Part II

Insurance

13 Loss distributions

Krzysztof Burnecki, Grzegorz Kukla, Rafal Weron

13.1 Introduction

14 Visualization of the risk process

Pawel Mista, Rafal Weron

14.1 Introduction

15 Approximation of ruin probability

Krzysztof Burnecki, Pawel Mista, Aleksander Weron

15.1 Introduction

16 Deductibles

*Krzysztof Burnecki, Joanna Nowicka-Zagrajek, Aleksander Weron, A.
Wyłomańska*

16.1 Introduction

17 Premium calculation

Krzysztof Burnecki, Joanna Nowicka-Zagrajek, W. Otto

17.1 Introduction

18 Premium calculation when independency and normality assumptions are relaxed

W. Otto

18.1 Introduction

19 Joint decisions on premiums, capital invested in insurance company, rate of return on that capital and reinsurance

W. Otto

19.1 Introduction

17219 *Joint decisions on premiums, capital invested in insurance company, rate of return on that ca*

20 Stable Levy motion approximation in collective risk theory

Hansjoerg Furrer, Zbigniew Michna, Aleksander Weron

20.1 Introduction

21 Diffusion approximations in risk theory

Zbigniew Michna

21.1 Introduction