Knock Out Power Options In Foreign Exchange Markets

Tome Eduardo Sicuaio[1] and Johan Tysk[2]

[1] Eduardo Mondlane University, Maputo , Mozambique
(E-mail: `tomeeduardosicuaio@uem.mz, tomeeduardosicuaio@gmail.com`)
[2] Uppsala University, Uppsala, Sweden
(E-mail: `johan.tysk@math.uu.se`)

Abstract. In recent years, the pressure of governments in maintaining currency parity has led to the break down of quite a few exchange rate mechanisms and has, thus, strengthened the need for companies, in particular, to make foreign exchange hedging decisions in order to avoid erosion of profit margins.
This paper deals with the pricing of Foreign exchange options. The Knock out options and the Power options will be treated in the sense that their payoffs are computed in closed form. We also combine the previous options in order to yield a new financial product. We find an explicit pricing formula for such financial product.
Additionally, a new product of First generation exotic options, called Knock out power options, is described and its payoff is computed and compared with the payoff of the Knock out options.
We have established explicit formulas for Symmetric and Asymmetric down and out power calls as well as for Symmetric and Asymmetric up and out power calls. To do this we used the Reflection Principle for Brownian motion.
The main difference between these instruments and standard Barrier options is that these instruments have a potentially higher discounted payoff. Therefore, these instruments can be used for speculation with power call options for an investor believing in an increasing exchange rate, and the put options for someone with the opposite view.
Keywords: Knock Out Options, Foreign Exchange Options, Power Options, Barrier Options, Symmetric Power Options, Asymmetric Power Options, First Generation Exotic Options and Second Generation Exotic options..

1 Introduction

Foreign exchange options are of great importance because nowadays are basic ingredients to international trading, they are many and each with its strategy for hedging.
Wytup [2] in his work describes the Power options whose payoff is just a vanilla option (call or put) raised to the power of n, that is,$[(X_T-K)^+]^n$, $[(K-X_T)^+]^n$ in the symmetric case; and each term of vanilla option(call or put) is raised to the power of n, that is,$(X_T^n - K^n)^+$, $(K^n - X_T^n)^+$ in the asymmetric case. Therefore, we encounter two kinds of Power options, the asymmetric and symmetric.

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

Furthermore, Wytup [2] also realizes that these options (Power options) are always equipped with high payoff compared to vanilla options. This, limits the risk of a short position even as the option premium for the holder.
This is one of the reasons that motivates speculators to invest in Power options once they request a high option premium. Yet, Wytup [2] and Tompkins [15] discuss hedging possibilities for Power options and they mentions that these options could be hedged using a combination of vanilla options with different strike prices.
Knock out options are Barrier options options whose payoff a Vanilla option if up to maturity time the underlying foreign exchange rate never hits the barrier agreed previously.
The holder of the Knock out option provides protection against the rising of the foreign exchange rate and according to this protection the holder has to pay a premium.
Lipton [10] discusses Barrier options for Foreign exchange options and give different forms of calculation of their payoff under risk neutral valuation.
Wytup [2] describes some of advantages and disadvantages of Knock out options. Ones of the advantages are: Knock out options are cheaper than Vanilla options, Knock out options provide a conditional protection against, for example, stronger USD/weaker SEK, Knock out options give a complete participation, for example, in a weaker USD/stronger SEK. Some of the disadvantages of Knock out options are: The exchange rate may hit the barrier before maturity time and another one is having to pay a premium.
One of the main aims in this Paper is to combine both options, Knock out options and Power options, and forming a new financial product in foreign exchange option which we will call it as Knock out power option. And we will compute its pricing function under risk neutral valuation using different methods.
The Paper consists of five sections, the first one describe foreign exchange options, and herein, we give an overview about foreign exchange options, the dynamics od foreign exchange rate are derived and is also described the pricing partial differential equation (known as Black-Scholes partial differential equation) for foreign exchange options.
In section two we describe the joint density of Brownian motion and its maximum and minimum. For this purpose we discuss the Reflection Principle in which we derive the main probability result that will be used in description of the joint density of Brownian motion and its maximum and minimum.
In the third section we describe the types of Foreign exchange options and special attention is given to Knock out options and Power options, and we compute the payoff under risk neutral valuation using different methods.
In the fourth section we describe our new financial product, Knock out power option, and also evaluate the payoff under risk neutral of the Down and out asymmetric power call option and Up and out symmetric power call option, using different methods. We also describe the sensitivity's analysis and we discuss the static hedging for Up and out power call option.
In the last section, the fifth, we describe the conclusions about our new financial product, Knock out power option.

2 Knock Out Power Options

Knock out options are Barrier options which pay a Power options payoff if up to maturity the underlying asset never reach the barriers. There are several types of Knock out power options but, we are only going to consider two kind of them, the Up and out power options and the Down and out power options. Suppose that the contractual parameters X_T is the underlying exchange rate, K the strike price, B the barrier and T the maturity time, however, the payoff functions are as shown below

$$\text{Down and Out}\begin{cases} \text{Asymmetric} & \begin{cases} (X_T^n - K^n)^+ \mathbf{1}_{\{X_t > B,\ t\in[0,T]\}} & \text{CALL} \\ (K^n - X_T^n)^+ \mathbf{1}_{\{X_t > B,\ t\in[0,T]\}} & \text{PUT} \end{cases} \\ \text{Symmetric} & \begin{cases} \left[(X_T - K)^+\right]^n \mathbf{1}_{\{X_t > B,\ t\in[0,T]\}} & \text{CALL} \\ \left[(K - X_T)^+\right]^n \mathbf{1}_{\{X_t > B,\ t\in[0,T]\}} & \text{PUT} \end{cases} \end{cases}$$

$$\text{Up and Out}\begin{cases} \text{Asymmetric} & \begin{cases} (X_T^n - K^n)^+ \mathbf{1}_{\{X_t < B,\ t\in[0,T]\}} & \text{CALL} \\ (K^n - X_T^n)^+ \mathbf{1}_{\{X_t < B,\ t\in[0,T]\}} & \text{PUT} \end{cases} \\ \text{Symmetric} & \begin{cases} \left[(X_T - K)^+\right]^n \mathbf{1}_{\{X_t < B,\ t\in[0,T]\}} & \text{CALL} \\ \left[(K - X_T)^+\right]^n \mathbf{1}_{\{X_t < B,\ t\in[0,T]\}} & \text{PUT} \end{cases} \end{cases}$$

The symbol S^+ denotes the positive parte of S, that is, $S^+ \equiv \max(S, 0)$.

The Up and out Asymmetric power call option,

it pays off if up to maturity time the underlying foreign exchange rate is between the exercise price K and the upper barrier B otherwise is zero, we can also say that the option is out of the money if the spot lies below the strike price K, at the money if the spot lies at exercise price and out of the money if the spot lies above the upper barrier B. On the other hand, for a Down and out Asymmetric power call option

it pays off if up to exercise date the underlying foreign exchange rate is above the lower barrier B, otherwise is zero, herein we can also say the option is in the money if the spot always lies above the lower barrier B and out of the money if the spot lies below the lower barrier B.

The Up and out symmetric power call option, it pays off if up to maturity time the underlying foreign exchange rate is between the exercise price K and the upper barrier B otherwise is zero, we can also say that the option is out of the money if the spot lies below the strike price K, at the money if the spot lies at exercise price and out of the money if the spot lies above the upper barrier B. On the other hand, for a Down and out symmetric power call option, it pays off if up to exercise date the underlying foreign exchange rate is above the lower barrier B, otherwise is zero, herein we can also say the option is in the

money if the spot always lies above the lower barrier B and out of the money if the spot lies below the lower barrier B.
We assume that all kind of Knock out Power options in foreign exchange options satisfy the same Black-Scholes partial differential equation.

$$\begin{cases} \frac{\partial F}{\partial t} + (r_d - r_f)x\frac{\partial F}{\partial x} + \frac{1}{2}\sigma_X^2 x^2 \frac{\partial^2 F}{\partial x^2} - r_d F = 0, \\ F(T,x) = \Phi(x). \end{cases}$$

3 Discounted payoff valuation

3.1 Up and out Asymmetric Power call option

Consider an European Up and out asymmetric power call option with exercise time T, strike price K and an up and out barrier B. We assume that the strike price K is lesser than the Up and out barrier B, because otherwise, the option must knock out in order to be in the money and hence could only pay off zero. Consider the spot foreign exchange rate

$$X_T = x \cdot e^{\left(r_d - r_f - \frac{1}{2}\sigma^2\right)T + \sigma \cdot \bar{W}},$$

where $\bar{W}$ is a Brownian motion with mean zero and variance one. Let us write the above foreign exchange rate as follows

$$X_T = x \cdot e^{\sigma W}, \tag{1}$$

where $W = \alpha T + \bar{W}, \quad \alpha = \frac{r_d - r_f - \frac{1}{2}\sigma^2}{\sigma}$ that is, W is a Brownian motion with mean αT and variance zero. Let

$$\bar{M}(T) = \max_{0 \le t \le T} W(t) \ \text{ and } \ \underline{M}(T) = \min_{0 \le t \le T} W(t)$$

be the maximum and minimum of W, respectively. Regarding the following identity

$$\begin{aligned} X_T > K \ \wedge \ \max_{0 \le t \le T} X_t < B &\equiv xe^{\sigma W(T)} > K \ \wedge \ \max_{0 \le t \le T} xe^{\sigma W(t)} < B \\ &\equiv W(T) > \frac{1}{\sigma} \ln \frac{K}{x} \ \wedge \ xe^{\sigma \max\limits_{0 \le t \le T} W(t)} < B \\ &\equiv W(T) > \frac{1}{\sigma} \ln \frac{K}{x} \ \wedge \ \max_{0 \le t \le T} W(t) < \frac{1}{\sigma} \ln \frac{B}{x} \\ &\equiv W(T) > k \ \wedge \ M(T) < b, \end{aligned}$$

where $k = \frac{1}{\sigma}\ln\frac{K}{x}$ and $b = \frac{1}{\sigma}\ln\frac{B}{x}$, the arbitrage free price for up and out power call option is given by

$$
\begin{aligned}
F(0,x) &= e^{-r_d T} E^{\bar{Q}}{}_{t,x}\left[\left(X_T^n - K^n\right)^+ \mathbf{1}_{\{\max\limits_{0\le t\le T} X_t < B\}}\right] \\
&= e^{-r_d T} E^{\bar{Q}}{}_{t,x}\left[\left(X_T^n - K^n\right) \mathbf{1}_{\{X_T > K \ \wedge \ \max\limits_{0\le t\le T} X_t < B\}}\right] \\
&= e^{-r_d T} E^{\bar{Q}}{}_{t,x}\left[\left(X_T^n - K^n\right) \mathbf{1}_{\{W(T) > k \ \wedge \ M(t) < b\}}\right] \\
&= e^{-r_d T} \int_k^b \int_{w^+}^b \left(x^n e^{n\sigma w} - K^n\right) \frac{2(2m-w)}{T\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2}\alpha^2 T - \frac{1}{2T}(2m-w)^2\right\} dm dw \\
&= e^{-r_d T - \frac{1}{2}\alpha^2 T} \int_k^b \left(x^n e^{n\sigma w} - K^n\right) \frac{1}{\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2T}(2m-w)^2\right\}\Bigg|_{w^+}^{b} dw \\
&= e^{-r_d T - \frac{1}{2}\alpha^2 T} \int_k^b \left(x^n e^{n\sigma w} - K^n\right) \frac{1}{\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2T}(2b-w)^2\right\} dw \\
&\quad - e^{-r_d T - \frac{1}{2}\alpha^2 T} \int_k^b \left(x^n e^{n\sigma w} - K^n\right) \frac{1}{\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2T}w^2\right\} dw \\
&= I_1 - I_2.
\end{aligned}
$$

Herein, we split up the integrals I_1 and I_2 as follows

$$
\begin{aligned}
I_1 &= e^{-r_d T - \frac{1}{2}\alpha^2 T} \int_k^b \left(x^n e^{n\sigma w} - K^n\right) \frac{1}{\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2T}(2b-w)^2\right\} dw \\
&= e^{-r_d T - \frac{1}{2}\alpha^2 T} x^n \int_k^b e^{n\sigma w} \frac{1}{\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2T}(2b-w)^2\right\} dw \\
&\quad - e^{-r_d T - \frac{1}{2}\alpha^2 T} K^n \int_k^b \frac{1}{\sqrt{2\pi T}} \exp\left\{\alpha w - \frac{1}{2T}(2b-w)^2\right\} dw \\
&= I_{11} - I_{12}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
I_2 &= e^{-r_dT-\frac{1}{2}\alpha^2T}\int\limits_k^b \left(x^n e^{n\sigma w} - K^n\right)\frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w - \frac{1}{2T}w^2\right\}dw \\
&= e^{-r_dT-\frac{1}{2}\alpha^2T}x^n\int\limits_k^b e^{n\sigma w}\frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w - \frac{1}{2T}w^2\right\}dw \\
&\quad -e^{-r_dT-\frac{1}{2}\alpha^2T}K^n\int\limits_k^b \frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w - \frac{1}{2T}w^2\right\}dw \\
&= I_{21} - I_{22}.
\end{aligned}
$$

Now, let us compute the closed form of each integrals $I_{11},\ I_{12},\ I_{21},\ I_{22}$ as follows

$$
\begin{aligned}
I_{11} &= e^{-r_dT-\frac{1}{2}\alpha^2T}x^n\int\limits_k^b e^{n\sigma w}\frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w - \frac{1}{2T}(2b-w)^2\right\}dw \\
&= e^{-r_dT-\frac{1}{2}\alpha^2T}x^n\int\limits_k^b \frac{1}{\sqrt{2\pi T}}\exp\left\{(n\sigma+\alpha)w - \frac{1}{2T}(2b-w)^2\right\}dw \\
&= e^{-r_dT+2b(n\sigma+\alpha)+n\sigma\alpha T+\frac{1}{2}n^2\sigma^2T}x^n\int\limits_k^b \frac{1}{\sqrt{2\pi T}}\exp\left\{-\frac{1}{2T}(w-(2b+(n\sigma+\alpha)T))^2\right\}dw \\
&= e^{-r_dT+2b(n\sigma+\alpha)+n\sigma\alpha T+\frac{1}{2}n^2\sigma^2T}x^n\int\limits_{\frac{k-(2b+(n\sigma+\alpha)T)}{\sqrt{T}}}^{\frac{b-(2b+(n\sigma+\alpha)T)}{\sqrt{T}}} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy \\
&= e^{-r_dT+2b(n\sigma+\alpha)+n\sigma\alpha T+\frac{1}{2}n^2\sigma^2T}x^n\left[N\left(\frac{b-(2b+(n\sigma+\alpha)T)}{\sqrt{T}}\right) - \right. \\
&\quad \left. N\left(\frac{k-(2b+(n\sigma+\alpha)T)}{\sqrt{T}}\right)\right] \\
&= \exp\left(-r_dT+n\Big(r_d-r_f+\frac{n-1}{2}\sigma^2\Big)T\right)\left(\frac{B}{x}\right)^{2\frac{r_d-r_f+(n-\frac{1}{2})\sigma^2}{\sigma^2}}x^n\Bigg[\\
&\quad N\left(\frac{\ln\frac{x}{B}-(r_d-r_f+(n-\frac{1}{2})\sigma^2)T}{\sigma\sqrt{T}}\right) \\
&\quad -N\left(\frac{\ln\frac{xK}{B^2}-(r_d-r_f+(n-\frac{1}{2})\sigma^2)T}{\sigma\sqrt{T}}\right)\Bigg].
\end{aligned}
$$

$$
\begin{aligned}
I_{12} &= e^{-r_dT-\frac{1}{2}\alpha^2T}K^n\int\limits_k^b\frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w-\frac{1}{2T}(2b-w)^2\right\}dw\\
&= e^{-r_dT-\frac{1}{2}\alpha^2T+2b\alpha+\frac{1}{2}\alpha^2T}K^n\int\limits_k^b\frac{1}{\sqrt{2\pi T}}\exp\left\{-\frac{1}{2T}(w-(2b+\alpha T))^2\right\}dw\\
&= e^{-r_dT+2b\alpha}K^n\int\limits_{\frac{k-(2b+\alpha T)}{\sqrt{T}}}^{\frac{b-(2b+\alpha T)}{\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy\\
&= e^{-r_dT+2b\alpha}K^n\left(N\left(\frac{b-(2b+\alpha T)}{\sqrt{T}}\right)-N\left(\frac{k-(2b+\alpha T)}{\sqrt{T}}\right)\right)\\
&= e^{-r_dT}\left(\frac{B}{x}\right)^{\frac{2(r_d-r_f-\frac{1}{2}\sigma^2)}{\sigma^2}}K^n\left(N\left(\frac{\ln\frac{x}{B}-(r_d-r_f-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right.\\
&\quad\left.-N\left(\frac{\ln\frac{xK}{B^2}+(r_d-r_f-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right).
\end{aligned}
$$

$$
\begin{aligned}
I_{21} &= e^{-r_dT-\frac{1}{2}\alpha^2T}x^n\int\limits_k^b e^{n\sigma w}\frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w-\frac{1}{2T}w^2\right\}dw\\
&= e^{-r_dT-\frac{1}{2}\alpha^2T}x^n\int\limits_k^b\frac{1}{\sqrt{2\pi T}}\exp\left\{(n\sigma+\alpha)w-\frac{1}{2T}w^2\right\}dw\\
&= e^{-r_dT-\frac{1}{2}\alpha^2T+\frac{1}{2}(n\sigma+\alpha)^2T}x^n\int\limits_k^b\frac{1}{\sqrt{2\pi T}}\exp\left\{-\frac{1}{2T}[w-(n\sigma+\alpha)T]^2\right\}dw\\
&= e^{-r_dT+n\sigma\alpha T+\frac{1}{2}n^2\sigma^2T}x^n\int\limits_{\frac{k-(n\sigma+\alpha)T}{\sqrt{T}}}^{\frac{b-(n\sigma+\alpha)T}{\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy\\
&= \exp\left\{-r_dT+n\Big(r_d-r_f-\frac{1}{2}\sigma^2\Big)T+\frac{1}{2}n^2\sigma^2T\right\}x^n\left[N\left(\frac{b-(n\sigma+\alpha)T}{\sqrt{T}}\right)\right.\\
&\quad\left.-N\left(\frac{k-(n\sigma+\alpha)T}{\sqrt{T}}\right)\right]\\
&= \exp\left\{-r_dT+n\Big(r_d-r_f+\frac{n-1}{2}\sigma^2\Big)T\right\}x^n\left[N\left(\frac{\ln\frac{B}{x}+\big(r_d-r_f+(n-\frac{1}{2})\sigma^2\big)T}{\sigma\sqrt{T}}\right)\right.\\
&\quad\left.-N\left(\frac{\ln\frac{K}{x}+\big(r_d-r_f+(n-\frac{1}{2})\sigma^2\big)T}{\sigma\sqrt{T}}\right)\right].
\end{aligned}
$$

$$
\begin{aligned}
I_{22} &= e^{-r_dT-\frac{1}{2}\alpha^2T}K^n\int\limits_k^b \frac{1}{\sqrt{2\pi T}}\exp\left\{\alpha w-\frac{1}{2T}w^2\right\}dw\\
&= e^{-r_dT-\frac{1}{2}\alpha^2T+\frac{1}{2}\alpha^2T}K^n\int\limits_k^b \frac{1}{\sqrt{2\pi T}}\exp\left\{-\frac{1}{2T}(w-T\alpha)^2\right\}dw\\
&= e^{-r_dT}K^n\int\limits_{\frac{k-T\alpha}{\sqrt{T}}}^{\frac{b-T\alpha}{\sqrt{T}}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}dy\\
&= e^{-r_dT}K^n\left[N\left(\frac{b-T\alpha}{\sqrt{T}}\right)-N\left(\frac{k-T\alpha}{\sqrt{T}}\right)\right]\\
&= e^{-r_dT}K^n\left[N\left(\frac{\ln\frac{B}{x}-\big(r_d-r_f-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)-N\left(\frac{\ln\frac{K}{x}-\big(r_d-r_f-\frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right)\right].
\end{aligned}
$$

Putting all together, that is, $F(0,x)=I_1-I_2=I_{11}-I_{12}-I_{21}+I_{22}$, we have

$$
\begin{aligned}
F(0,x) = e^{-r_dT}\Bigg\{ &-e^{n\left(r_d-r_f+\frac{n-1}{2}\sigma^2\right)T}\left(\frac{B}{x}\right)^{2\frac{r_d-r_f+(n-\frac{1}{2})\sigma^2}{\sigma^2}}x^n\left[N\left(\frac{\ln\frac{x}{B}-\delta_n}{\sigma\sqrt{T}}\right)\right.\\
&\left.-N\left(\frac{\ln\frac{xK}{B^2}-\delta_n}{\sigma\sqrt{T}}\right)\right]+\left(\frac{B}{x}\right)^{\frac{2(r_d-r_f-\frac{1}{2}\sigma^2)}{\sigma^2}}K^n\left[N\left(\frac{\ln\frac{x}{B}-\delta}{\sigma\sqrt{T}}\right)\right.\\
&\left.-N\left(\frac{\ln\frac{xK}{B^2}-\delta}{\sigma\sqrt{T}}\right)\right]+\exp\left\{n\Big(r_d-r_f+\frac{n-1}{2}\sigma^2\Big)T\right\}x^n\\
&\cdot\left[N\left(\frac{\ln\frac{B}{x}-\delta_n}{\sigma\sqrt{T}}\right)-N\left(\frac{\ln\frac{K}{x}-\delta_n}{\sigma\sqrt{T}}\right)\right]\\
&-K^n\left[N\left(\frac{\ln\frac{B}{x}-\delta}{\sigma\sqrt{T}}\right)-N\left(\frac{\ln\frac{K}{x}-\delta}{\sigma\sqrt{T}}\right)\right]\Bigg\}.
\end{aligned}
$$

where

$$\delta=\big(r_d-r_f-\frac{1}{2}\sigma^2)T \text{ and } \delta_n=\Big(r_d-r_f+(n-\frac{1}{2})\sigma^2\Big)T.$$

This result can be written using financial variables as follows:

Proposition 1. *The price of an Up and out Asymmetric power call option with an upper barrier B, exercise price K and maturity time T, regarding that*

$K < B$, is given by the formula

$$
\begin{aligned}
C_{UAOAPO}(t, x, T, K, B) = C_{APO}(t, x, T, K) - C_{APO}(t, x, T, B) \\
- \left(\frac{B}{x}\right)^{2\frac{r_d - r_f}{\sigma^2} - 1} \Bigg[C_{APO}\Big(t, \frac{B^2}{x}, T, K\Big) - \\
C_{APO}\Big(t, \frac{B^2}{x}, T, B\Big) \Bigg] - \\
(B^n - K^n) \Bigg[V^{CB}(t, x, T, B) - \\
\left(\frac{B}{x}\right)^{2\frac{r_d - r_f}{\sigma^2} - 1} V^{CB}\Big(t, \frac{B^2}{x}, T, B\Big) \Bigg],
\end{aligned}
$$

where $V^{CB}(t, x, T, K)$ is the price of a call bet which pays one unit of domestic currency if the foreign exchange rate at maturity is above K, and otherwise nothing.

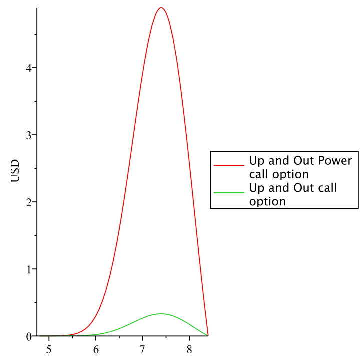


Fig. 1. Up and out Asymmetric power call options using (Volatility) $\sigma = 0.0853$, (Interest rates) $r_{USD} = 0.08$, $r_{SEK} = 0.12$, (Maturity time) $T = 1$year, (Strike price) $K = 7$, (Upper barrier) $B = 8.4$, $n = 2$.

It is clearly seen from Figure 1 that the payoff of Up and out power call option remain higher than the payoff of the Up and out call option. This fact can obviously motivate the Speculators to invest in a product with higher payoff.

3.2 Up and out Symmetric Power call option

Consider an European symmetric power call option with exercise time T, strike price K and an up and out barrier B. We assume that the strike price K is lesser than the up and out barrier B, because otherwise, the option must knock

out in order to be in the money and hence could only pay off zero.
Following a similar procedure as used in the previous payoff valuation and applying Newton's binomial theorem we have the discounted payoff for an Up and out symmetric power call option stated by the following proposition.

Proposition 2. *The price of an Up and out symmetric power call option with an upper barrier B, exercise price K and maturity time T, regarding that $K < B$, is given by the formula*

$$F(0,x) = e^{-r_d T}\sum_{j=0}^{n}\binom{n}{j}(-K)^j x^{n-j}\exp\left(\left[r_d - r_f + \Big(\frac{n-j-1}{2}\Big)\sigma^2\right](n-j)T\right)$$
$$\left\{\Big(\frac{B}{x}\Big)^{\frac{2\cdot\eta}{\sigma^2}}\cdot\left[N\left(\frac{\ln\frac{x}{B}-\eta\cdot T}{\sigma\sqrt{T}}\right) - N\left(\frac{\ln\frac{xK}{B^2}-\eta\cdot T}{\sigma\sqrt{T}}\right)\right]\right.$$
$$\left.-\left[N\left(\frac{\ln\frac{B}{x}-\eta\cdot T}{\sigma\sqrt{T}}\right) - N\left(\frac{\ln\frac{K}{x}-\eta\cdot T}{\sigma\sqrt{T}}\right)\right]\right\},$$

where

$$\eta = r_d - r_f + (n-j-\frac{1}{2})\sigma^2.$$

3.3 Down and out Symmetric Power call option

Consider an European symmetric power call option with exercise time T, strike price K and a down and out barrier B. We assume that the strike price K is greater than the down and out barrier B, because otherwise, the option must knock out in order to be in the money and hence could only pay off zero.
The mathematical model for the European down and out symmetric power call its payoff function is stated in the following proposition:

Proposition 3. *The price of a Down and out symmetric power call option with a lower barrier B, exercise price K and maturity time T, regarding that $K > B$, is given by the formula*

$$C_{DOSPO}(t,x,T,K,B) = C_{SPO}(t,x,T,K) - \Big(\frac{x}{B}\Big)^{2\alpha} C_{SPO}\Big(t,\frac{B^2}{x},T,K\Big), \quad (2)$$

where $C_{DOSPO}(t,x,T,K,B)$ is the price of a Down and out symmetric power call option and $C_{SPO}(t,x,T,K)$ is the price of Symmetric power call option under risk neutral valuation.

3.4 Down and out Asymmetric Power call option

Consider an European asymmetric power call option with exercise time T, strike price K and a down and out barrier B. We assume that the strike price K is greater than the down and out barrier B, because otherwise, the option must knock out in order to be in the money and hence could only pay off zero.
The mathematical model for the European down and out symmetric power call is

Proposition 4. *The price of a Down and out asymmetric power call option with a lower barrier B, exercise price K and maturity time T, regarding that $K > B$, is given by the formula*

$$C_{DAOAPO}(t, x, T, K, B) = C_{APO}(t, x, T, K) - \Big(\frac{x}{B}\Big)^{2\alpha} C_{APO}\Big(t, \frac{B^2}{x}, T, K\Big), \tag{3}$$

where $C_{DAOAPO}(t, x, T, K, B)$ is the price of a Down and out asymmetric power call option and $C_{APO}(t, x, T, K)$ is the price of Asymmetric Power Call Option under risk neutral valuation.

Fig. 2. Discounted payoff of Down and out asymmetric power call and Down and out call (Strike price) $K = 7$, (Lower barrier) $B = 4.8$, (Interest rates) $r_{USD} = 0.08$, $r_{SEK} = 0.12$, (Volatility) $\sigma = 0.0853$, (Maturity time) $T = 1$year, $n = 2$

In Figure 2, it is clearly seen that the Down and out asymmetric power call has a higher discounted payoff function than the simple Down and out call.

4 Hedging with Knock out Power Options

In this section we will discuss the static hedging of the Up and out asymmetric power call option by creating it synthetically from power call options and power put options. To do so, we will follow the sophisticated case discussed in Lipton [10] for up and out call option. Maruhn [11] has also described the static hedging for up and out call option.

A static hedge is a strategy which is fixed and is created onetime to hedge a prevailing position or option. It is not adjusted at all, once created, in contrast with a dynamic hedge. An overview of static hedge is to provide an exact and necessary protection so far and in addition including the option maturity date.

One of the reasons of hedging is to be immune to unpleasant surprises, faced by the portfolio, such as sharp increment, for example, in price of foreign exchange rate.

Up and out power call option and Down and out power put option have discontinuity point in the boundary, this makes their portfolio difficulty for hedging.

Set $n = 2$ and consider the payoff of an Up and out asymmetric call option,

$$(X_T^2 - K^2)^+ \mathbf{1}_{\{X_t < B,\ t \in [0,T]\}}. \tag{4}$$

The payoff (4) can be written as follows,

$$(X_T^2 - K^2)\mathbf{1}_{\{X_T > K,\ X_t < B,\ t \in [0,T]\}}. \tag{5}$$

Furthermore

$$X_T^2 - K^2 - (X_T - K)^2 = 2K(X_T - K). \tag{6}$$

Though, we can multiply the above relation (6) by $\mathbf{1}_{\{X_T > K,\ X_t < B,\ t \in [0,T]\}}$ and we get

$$(X_T^2 - K^2)^+ \mathbf{1}_{\{X_t < B,\ t \in [0,T]\}} = [(X_T - K)^+]^2 \mathbf{1}_{\{X_t < B,\ t \in [0,T]\}} + 2K(X_T - K)^+ \mathbf{1}_{\{X_t < B,\ t \in [0,T]\}}.$$

This, equality shows that an Up and out asymmetric power call option can be hedged by purchasing $2K$ units of standard Up and out call option and buying a Symmetric up and out power option.

5 Conclusion

We have established explicit formulas for Symmetric and Asymmetric down and out power calls as well as for Symmetric and Asymmetric up and out power calls. To do this we used the Reflection Principle for Brownian motion.
The main difference between these instruments and standard Barrier options is that these instruments have a potentially higher discounted payoff.
Therefore, these instruments can be used for speculation with power call options for an investor believing in an increasing exchange rate, and the put options for someone with the opposite view.

References

1. Bjork, Tomas. *Arbitrage theory in continuous time*, Oxford University Press, New York, 2009.
2. Wystup, Uwe. *FX Options and Strutured Products*, Wiley, British Library, 2007.
3. Glasserman, Paul. *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2004.
4. Shamah, Shani. *A Currency options primer*, Wiley, 2004.
5. Shreve, S. E. *Stochastic Calculus for Finance I+II*, Springer,2004.
6. Kwok, Yue-Kuen. *Mathematical Model of Financial Derivatives*, Springer,2004.
7. Jiang, Lishang. *Mathematical Modeling and Methods of Option Pricing*, World Scientific, 2005.
8. Strauss, Walter A. *Partial Differential Equations An Introduction*, John Wiley & Sons, Inc., 1992.
9. Hirsa, Ali. Neftci, Salih N. *An Introduction to the Mathematics of Financial Derivatives*, Academic Press, 2013.
10. Lipton, Alexander. *Mathematical Methods for Foreign Exchange - A Financial Engineer's Approach*, World Scientific, 2001.
11. Maruhn, Jan H. *Robust Static Super-Replication of Barrier Options*, Walter de Gruyter, 2009.
12. Hull, John C. *Options, Futures, and Other Derivatives*, Pearson, 2012.
13. Moosa, Imad A. *International Financial Operations, Arbitrage, Hedging, Speculation and Investment*, Palgrave Macmillan, 2003.
14. Madura, Jeff. Fox, Roland. *International Financial Management*, South Western-Cengage learning, 2011.
15. Tompkins, Robert G. *Power options: hedging nonlinear risks*, The journal of risks, Austria, 1999.
16. Khoshnevisan, Davar. *Probability*, American Mathematical Society, 2007.

Information Hidden in Regional Input-Output Tables[1]

Jaroslav Sixta[1], Jakub Fischer[2], and Kristyna Vltavska[3]

[1] University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3. Czech Republic
(E-mail: sixta@vse.cz)
[2] University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3. Czech Republic
(E-mail: fischerj@vse.cz)
[3] University of Economics in Prague, Dep. of Economic Statistics, Nam W. Churchilla 4, Prague 3. Czech Republic
(E-mail: kristyna.vltavska@vse.cz)

Abstract. Regional input-output tables offer an interesting and valuable data source. They can be used for lots of studies ranging from forecasting to detailed structural analysis. A pity is that they are not used very often since they are difficult to compile. Usually, official statistical authorities do not publish regional input-output with some exceptions (like the U.S. Bureau of Economic Analysis). It means that pure regional input-output tables are very scarce and regional input-output tables are usually compiled for groups of countries. We demonstrate how these tables can be compiled and used for relatively small regions of the Czech Republic, country with about 10 million inhabitants. Our regional input-output tables are product by product tables with 82 rows and columns and they are prepared for all 14 regions (NUTS 3 level) of the Czech Republic. These regional input-output tables combine officially published regional accounts and national symmetric input-output tables. The key issue of their construction lies in the definition of statistical unit (local kind of activity unit) and its decomposition between regions. The paper presents both brief description of the methodology and the example of forecasting issues based regional input-output tables. The case of dependency between the regions is deeply discussed and main issues are tackled and explained.
Keywords: Regional, Input-Output Tables

1 Introduction

Regional input-output tables represent a subject of very often discussed topic. This issue was deeply studied from several perspectives, mainly from theoretical point of view. Practical examples of usage of regional input-output tables are

[1] This paper is prepared under the support of the project "Regional estimates of gross domestic product based on the expenditure approach" of the Czech Science Foundation, project No 13-15771S.

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

very scarce. Availability of regional models respecting specific structure of a particular region cannot be easily found. Under the heading regional input-output table is usually described very broad region. Then region usually represents group of countries or states rather than districts and regions of one country. The importance of regional input-output tables is clear when discussing regional specifics. Regional analysis respecting regional economic relationships always provides better results than when using country averages. Our model comprises 14 regions of the Czech Republic described by symmetric input-output tables. It means that the tables are completely similar to national symmetric input-output tables. They are compiled on the level of CPA (Classification of Products) two digits, 82 products.

When preparing regional analyses, the lack of information is very often present. Input-output analysis based on regional input-output tables can be used for preparing plans for regions suffering with high unemployment and low demand, planning infrastructure, tax incentives for investors etc. Official statistics offer standard statistical data mainly covering regional employment and unemployment, gross value added, investments but it seems not enough for many purposes.

2 Methodology and data sources

Regional input-output tables (RIOTs) are data demanding. They have to base on the combination of different data sources. In our case, aggregates and totals are taken over from national accounts and regional accounts officially published by the Czech Statistical Office. These data sources do not cover complete production approach because only value added is used. From the expenditure approach only gross capital formation (GFC) is usually published in the EU (see Kahoun and Sixta, 2013). Officially published data sources have to be combined with other information from social statistics, trade and transport statistics. The following Table 1 describes key data sources.

Table 1 Key data sources for compilation regional input-output tables

	Name	Use
1	Regional Accounts	Structure of value added that represents a framework for the region, additional data sources for estimates of output. GFC
2	Symmetric input-output tables	Technical coefficients for intermediate consumption matrices
3	Transport statistics	National GDP, regional GDP, distance between regions
4	External estimates of expenditure approach based on previous research	Households (see Musil and Kramulová, 2013), government and non-profit institutions consumption

Source: Own elaboration

The correct combination of data sources is crucial when compiling RIOTs and consistency issues have to be discussed. The problem lies in standard procedures in modern regional and national accounts since the one of the most preferred method is top-down approach. It means that final national accounts aggregates are distributed into the regions in line with auxiliary indicators like wages, number of workers etc.

It is easier to distribute gross value added than output. It is due to the links between statistical units, local kind of activity units (LKAUs). It can be illustrated on the case of electricity company with many local units (power plants) and headquarter in the capital city. In many cases, LKAUs are usually not correctly defined as it is assumed in the theory. Then the case with electricity company provides interesting findings. The distribution of company value added usually done by wages can be easily justified. It means that the rate of profit (operating surplus) is fixed proportionally to wages. However, the output of company's headquarter is not production of electricity, it is a production of accounting, legal and management services. The problem is that there cannot be identified transaction between regional local kind of activity units and headquarter. Company's regional units do not pay both externally and internally for management services to headquarter. There are two possibilities how to solve it. The first consists in imputation of output (product management etc.) of headquarter and intermediate consumption of local units. But it means that total output will be lower than the sum of regional output. We selected the second option that consist in consistency between regional output and total output in national accounts. It based on the assumption that the producer in the capital city produces main products (electricity) with respect to its share. The problem of the second option lies in the fact that the production may remain without appropriate inputs. The cost of headquarter in the capital are composed from renting services, accounting services, advertisement costs etc. Coal and gas are consumed in the regional power plants that actually produce electricity. This issue needs correct interpretation. Electricity company produces electricity in all units including both headquarters and local branches and this actually represent the product that buyer buys. The output created by headquarter has a nature similar to margin, added value for the provision of the business. From this perspective, the most important is to associate costs to obtain reliable regional intermediate consumption matrices. We consider the quality of output measurement as secondary. It means that if the price of coal is raised, the price of final product (electricity) will be raised, as well. This will affect both regional power plants and headquarters in the capital. And this effect will work the same irrespective of the imputation of additional output.

3 Regional Input-Output Tables as a Source of Information

Regional input-output tables were constructed for 2011 in line with available data sources in ESA 1995 methodology. They provide lots of information about regional structures of output, intermediates and final use. The main reason for their construction lies in the description of regional specifics. Overall indicators constructed for total economy are not suitable for regional modeling. Even in smaller country, the Czech Republic, regional differences can be important. Mostly the capital city of Prague differs significantly from other regions.
Table 1 shows the structure of regional output according to CPA classification. The main share of output is produced in the capital city of Prague (24.4%) and in the Středočeský region (12.3%). From the products point of view the main portion of services is produced in the capital city of Prague and in the Jihomoravský region where the second largest Czech city takes place. It is caused e.g. by the concentration of the biggest companies and higher education institutions with all the research centers.

	A	**B+C+D+E**	**F**	**G+H+I**	**J**
Pha (CZ010)	3.5	9.5	25.7	32.7	61.8
Stc (CZ020)	15.4	16.3	9.6	11.7	4.0
Jhc (CZ031)	10.3	5.0	5.6	4.8	2.1
Plz (CZ032)	7.9	5.4	4.3	4.4	2.2
Kar (CZ041)	2.1	1.6	1.8	2.0	0.5
Ust (CZ042)	5.1	9.0	6.6	5.3	3.5
Lib (CZ051)	2.2	3.7	3.0	2.7	1.2
Krh (CZ052)	6.9	5.2	3.6	3.8	2.6
Par (CZ053)	6.4	6.3	3.6	3.4	3.2
Vys (CZ063)	9.3	4.8	4.7	3.1	1.1
Jhm (CZ064)	12.3	8.7	13.6	9.7	9.1
Olm (CZ071)	6.6	4.2	4.4	4.3	1.7
Zln (CZ072)	5.0	5.7	4.6	3.7	1.5
Mrs (CZ080)	6.9	14.4	8.9	8.5	5.5

	K	**L**	**M+N**	**O+P+Q**	**R+S+T+U**
Pha (CZ010)	67.1	38.6	48.9	22.6	32.9
Stc (CZ020)	2.9	9.8	6.9	9.1	9.1
Jhc (CZ031)	2.4	4.0	3.0	6.3	4.5
Plz (CZ032)	2.5	3.7	3.6	4.9	4.1

Kar (CZ041)	0.9	1.9	1.3	3.5	2.5
Ust (CZ042)	2.2	4.9	4.2	7.1	5.0
Lib (CZ051)	2.0	2.8	1.9	3.6	3.1
Krh (CZ052)	2.4	3.1	2.5	5.0	4.1
Par (CZ053)	2.9	2.9	2.5	4.3	3.0
Vys (CZ063)	1.1	3.3	1.8	3.9	2.3
Jhm (CZ064)	6.4	10.2	10.4	9.7	10.9
Olm (CZ071)	1.9	3.5	2.9	5.8	4.5
Zln (CZ072)	1.5	3.5	3.1	4.5	5.4
Mrs (CZ080)	3.7	7.7	7.1	9.9	8.7

Table 1. The regional structure of output

Note: A - Agriculture, forestry and fishing, B - Mining and quarrying, C – Manufacturing, D - Electricity, gas, steam and air conditioning supply, E - Water supply; sewerage, waste management and remediation activities, F – Construction, Services: G - Wholesale and retail trade; repair of motor vehicles and motorcycles, H - Transportation and storage, I - Accommodation and food service activities, J - Information and communication, K - Financial and insurance activities, L - Real estate activities, M - Professional, scientific and technical activities, N - Administrative and support service activities, O - Public administration and defense; compulsory social security, P – Education, Q - Human health and social work activities, R - Arts, entertainment and recreation, S - Other service activities, T - Activities of households as employers and producers for own use.

Note: Pha – the capital city of Prague, Stc – Středočeský region, Jhc – Jihočeský region, Plz – Plzeňský region, Kar – Karlovarský region, Ust – Ústecký region, Lib – Liberecký region, Krh – Královehradecký region, Par – Pardubický region, Vys – Vyšočina region, Jhm – Jihomoravský region, Olm – Olomoucký region, Zln – Zlínský region, Mrs – Moravskoslezský region

Different information is obtained when analyzing the inputs to the production process. Inputs into production are recorded in intermediate consumption. It is clear that different structure of industries requires different inputs. Therefore regional input-output tables can be also used for the modeling of impact of administrative decisions. E.g. the impact of the increased taxes on selected products is different in the regions in line with the intermediate consumption. Figure 1 describes the share of intermediate consumption of each product (CPA) on the total regional output, called input coefficient. Again, one can see that the capital city of Prague differs among the rest of the regions. In the capital city of Prague, the highest input coefficients refers to the products connected trade, transportation, accommodation and catering. The most of intermediate consumption is formed by services. Other regions have the highest input coefficients of products mining, quarrying and manufacturing intermediate consumption on the regional output. When we have a closer look on agriculture product the main portion of intermediate consumption is found in the central part of the Czech Republic (Vysočina Region 3.4%, Královehradecký region

2.2%). Real estate acvities (CPA L) represents one of the most important part of the economy. The main portions of intermediate consumption are found in the capital city of Prague (5.2%) and in Jihočeský region (2.6%) with the second biggest city of the country.

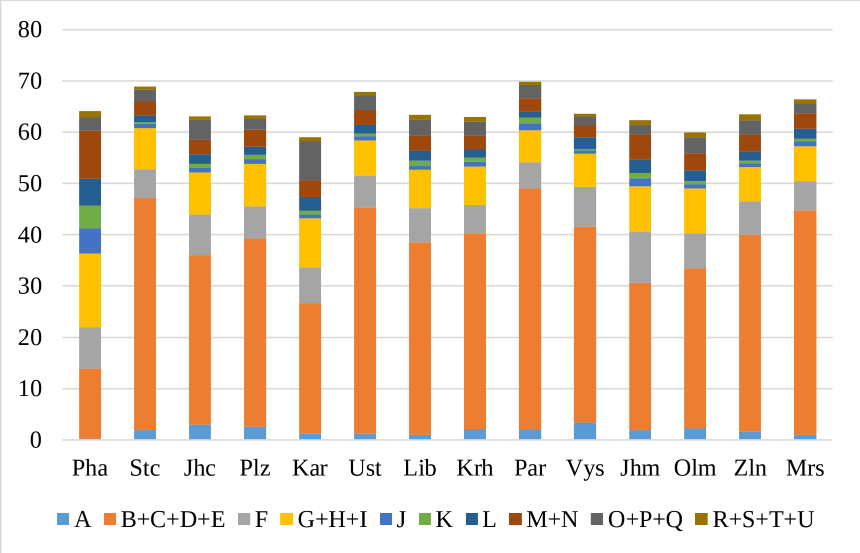


Fig. 1. The regional structure of intermediate consumption, %

The structure of the final use refers mainly to the wealth of the region. The structure of household consumption reflects the population habits and price relations. The quality of non-market services is described by the government consumption. Figure 2 shows the share of the final use of each product in the region on regional output. This figure again proves significant differences between the capital city of Prague and the rest of regions. The allocation of the final use on the total output mainly lies in services. The highest share of final use on regional output is observed in Karlovarský region. It is connected with low production in the region and low possibilities of export, see table 2. Even the output and production possibilities differ significantly, the consumption of population cannot be dramatically different. It is connected with very high unemployment rate (8.5%) in comparison with overall rate of 6.7% in 2011.

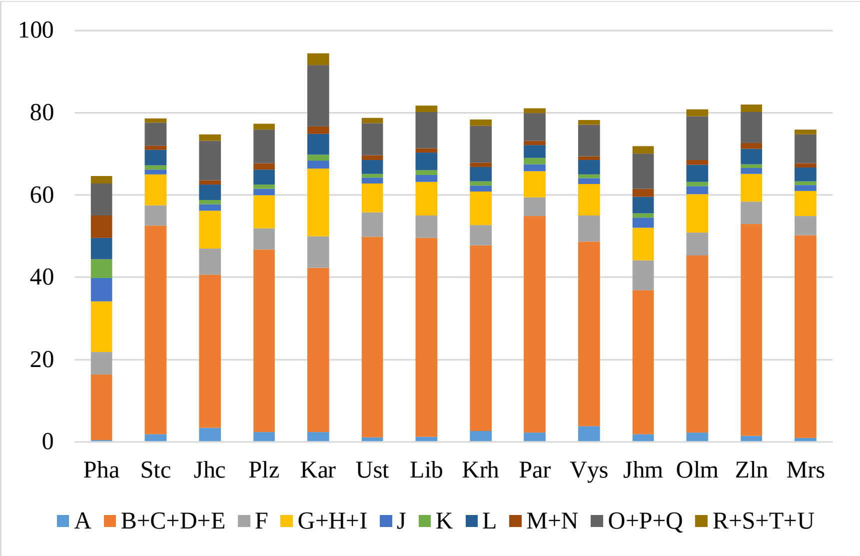


Fig. 2. The regional structure of final use, %

The exporting performance of the region is represented by the net export (export less import). Regional net export covers both interregional export and import and foreign trade. The sum of regional imports and exports is therefore higher than the figure for total economy. The resulting net export is the same. In six regions (Jihočeský, Karlovarský, Ústecký, Liberecký, Jihomoravský and Olomoucký) the import is higher than export so we can classified these regions as importers. The rest of the regions are exporters. It can be described for example by different product structure of each region and on the dissimilar structure of regional gross value added. Among all, the highest addition to the net export of the Czech Republic (6 bil EUR) is observed in the capital city of Prague (7 bil EUR). It is mostly connected with the headquarters of Czech companies. The headquarter has usually very high value added (high wages of top managers and the office staff). Besides the capital city of Prague, in Zlínský region and Moravskoslezský region, important companies connected with automotive industry are located. These companies usually export most of their output.

Czech Republic	5 997
Pha	7 449
Stc	240
Jhc	-149
Plz	127
Kar	-550

Ust	-990
Lib	-297
Krh	176
Par	86
Vys	29
Jhm	-329
Olm	-713
Zln	556
Mrs	361

Table 2. The regional net export, 2011, mil EUR

Conclusions

Regional input-output tables provide lots of information about particular regions. Though they are quite difficult to compile, we illustrated that they can be prepared for the regions of relatively small country as the Czech Republic. Presented data come from our original research and the methodology is described in Miller, Blair (2009). The most valuable information are connected with regional structure of output, different types of intermediates and the structure of final use. The position of the region can be also discussed from the perspectives of net export.

The results proves that the capital City of Prague differs from the rest of the Czech regions. There is the highest share of services on the total output. There is the highest input coefficients refers to the products connected with trade, transportation, accommodation and catering. The most of intermediate consumption is formed by services and the net export creates 7 bil EUR.

References

1. European Union (2013). Manual on regional accounts methods. Luxembourg, 2013.
2. EUROSTAT (1996). European System of Accounts – ESA 1995. Luxembourg, Office for Official Publications of the European Communities, 1996.
3. J. Kahoun and J. Sixta. Regional GDP Compilation: Production, Income and Expenditure Approach, Statistika, 93, 4, 24{36, 2013.
4. J. Kramulova and P. Musil. Experimental Estimate of Components of Expenditure Approach to Regional GDP in the Czech Republic, Politická ekonomie, 61, 6, 814{833, 2013.
5. R. E. Miller and P.D. Blair. Input-Output Analysis: Foundations and Extensions. 2009
6. J. Sixta. Development of Input-Output Tables in the Czech Republic, Statistika, 93, 2, 4{14, 2013.
7. J. Sixta, K. Vltavská and J. Fischer. The Development of Gross Domestic Product in the Czech Republic and Slovakia between 1970 and 1989, Ekonomický časopis, 61, 6, 549{562, 2013.

Comparing estimates of healthy life expectancy in Europe from WHO and from first exit time theory

Christos H. Skiadas[1] and Konstantinos N. Zafeiris[2]

[1] ManLab, Department of Production Engineering and Management, Technical University of Crete, Chania, Greece. (e-mail: skiadas@cmsim.net)
[2] Laboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace, Komotini, Greece. (e-mail: kzafiris@he.duth.gr)

Abstract: The first exit time theory's HLEB3 and the WHO estimations of Healthy life Expectancy (HALE) are compared in this paper. The application of first exit time theory was based on the abridged life tables published by WHO, which were expanded to full life tables with a non linear regression procedure. Findings indicate that the two methods are in accordance concerning the estimation of healthy life expectancy.

Keywords: HALE, HLEB3, WHO, First exit time theory.

1 Introduction

1.1 The World Health organization approach

According to the methodology developed by the World Health Organization (WHO) the healthy life expectancy (HALE) is defined as the average number of years a person is expected to live in "full health"[1]. The computation of HALE is based on Sullivan's method [21] in which population data on health and disability are combined in a life table [25]. Quite briefly, in this method the years lost due to disability (YLD) are estimated across a comprehensive set of disease and injury causes (see [24] [23]). Then, the per capita fraction of YLD for all causes is calculated for every age group, sex and country, after adjusting for independent comorbidity. Based on that fraction the lost years of healthy life are calculated for each age group and the Healthy Life Expectancy at age x is the sum of healthy life years from the age x up to the open ended interval of the life table divided by the survivors in each age x ([25] [24] [11] [9]). The new estimates of WHO for years 2000-2012, are based on the results of the Global Burden of Disease (GBD; [6] [10] [11]) study of 2010 after the revision of existing or the application of new methods for dealing with comorbidity. In that way they are not directly comparable with the WHO estimates of HALE for previous years [25]. However, as WHO notes, several limitations exist in the aforementioned method, because of the lack of reliable data on mortality and morbidity and of the comparability of self-reported data from health interviews and the measurement of health-state preferences for such self-reporting[1].

[1] http://apps.who.int/gho/indicatorregistry/App_Main/view_indicator.aspx?iid=66

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

An example of such calculations is found in Salomon et al. [14], which used three sets of data from the GBD 2010 study: estimates of the age specific mortality rates, of the prevalence of 1160 sequelae (pathological conditions resulting from a disease) by age, sex, year and country and the relevant disability weights. In order for the disability weights to be estimated, the 1160 sequelae were mapped into 220 unique health states and represent health loss, measured in a scale from 0 (equivalent to ideal health) to 1 (equivalent to death). They were measured on the basis of household surveys in five countries (Bangladesh, Indonesia, Peru, Tanzania and the USA) and open access web based surveys relying on simple paired comparison questions [13]. In such questions, the respondents were presented with two individuals in different health states described in lay language and asked which person they regarded as healthier. Then probit regression analysis was used in order to translate these paired responses into estimates of health on a continuous scale. Additional questions were used in the web surveys, in order for example, to compare the health benefits of different life-saving or disease prevention programs etc. This methodology is even more complicated as for the estimation of age specific rates a variety of sources was used along the official death registration data, such as household surveys, censuses. The same was the case for the prevalence of the sequalae [23]. Afterwards, a Monte Carlo simulation approach was used for computing the average health of individuals in a population within an age interval by taking into consideration the prevalences for all sequelae and their disability weights accounting also for comorbidity. Finally, the Sullivan method was applied [21]. Obviously enough, this is a rather complicated procedure, possibly subject to sample or other types of errors, as people from different cultures and socio-economic background may have different views, perceptions and experiences about health and disability. Because of that we proposed an alternative method.

1.2 The First Exit Time Theory approach

The First Exit Time Theory ([7] [18] [19] [20] [17]) is based on the idea that if the mortality of a population can be modeled on the basis of death and population data, then its health status can be modeled too with the same data. The implementation of this approach comes from the general theory of dynamic models for modeling human life introduced by Jansen and Skiadas [7]. In that way, while the health of an individual changes over time according to a stochastic process, their death comes when health falls below a limit, or a barrier as it is called in the first exit time theory. Then the problem is to find the distribution of the first exit time of a diffusion process expressing the health state of a person from a barrier [15].

It is proven [15] that the death density distribution g(x) or the death distribution d(x) in a life table can be modeled as:

$$d(x) = k(l + (c-1)(bx)^c)(x)^{-3/2} e^{-\frac{(l-(bx)^c)^2}{2x}} \quad (1)$$

where x is the age and k, l, c, b parameters which need to be estimated. In fact a non linear regression model is applied on the dx distribution of a full life table and then the Health State Function H(x) is estimated as:

$$ABS\left(-2xln\frac{d_{(x)}\sqrt{x^3}}{k}\right)^{\frac{1}{2}} \quad (2)$$

where k is estimated as k=max($d_{(x)}\sqrt{x^3}$) [16].

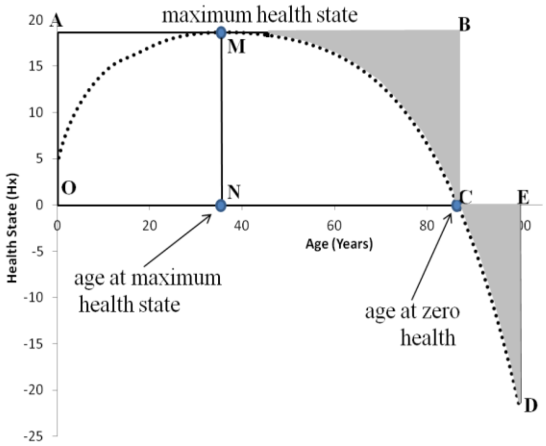


Fig 1. The health state function of the population up to the zero health age.

The Health State Function (HSF) has the form seen in Figure 1, in which health increases up to one point and decreases afterwards in order to be 0 in a time point of human life cycle located at the older ages (age at zero health). In fact the first part of HSF within the rectangle AMNO describes the phase of development during the human life cycle in which at a particular age the "maximum health state" of an organism is observed [17]. This point corresponds to the maximum vitality of that organism. The white area within the rectangle MBCN represents the deterioration phase of human health until its zero point. If no-deterioration mechanism was present, or the repairing mechanism of human body was perfect during that phase, then the health state would continue following the straight line AMB which is parallel to the x axis. This is not the case of course and that leads to the gradual disruption of human health. The problem is how to estimate the "lost life years" during the deterioration phase of the human life cycle. If THD_{ideal} is the ideal total dynamics of the population a geometric solution can be given as:

$$LHLY1 = \lambda \frac{OABC}{THD_{ideal}} \frac{THD_{ideal}}{MBCM} = \lambda \frac{OABC}{MBCM}$$

where λ is a parameter expressing years and should be estimated for every case and MBCM the grey area of the rectangle MNCB. It was found that for purposes of multiple comparison of countries λ could be set to be 1 year.

However, the above formula has to be expanded further if the health state of the people lived beyond the age at zero health is taken into consideration. In fact these people contributed to the health state of the population and for the sake of visualization they are represented in the area ECD. Then the equation above can be expanded in for a new estimation to be named LHLY3 to be calculated as:

$$LHLY3 = \lambda \frac{OABC + ECD}{MBCM} \quad (3)$$

Based on the last equation the healthy life expectancy can be simply calculated as: (life expectancy at birth)-LHLY3.

The scope of this paper is to make a comprehensive assessment of the healthy life expectancy (HLEB3) in the European countries estimated by the First Exit Time Theory in comparison to the estimations of the World Health Organization.

2 Data and methods

The life table data published by the World Health Organization (WHO, http://apps.who.int/gho/data) used for the application of the First Exit Time Theory. These life tables are in an abridged form and contain information for the age groups <1, 1-4 and for 5-years age intervals up to the age 100 which corresponds to the open-ended one.

However, the First Exit Time Theory is applied on full life tables and in that way the available life tables should be unabridged. Several procedures were used for that purpose; but in general it was found that many problems occurred during their appliance. For example the UNABR application[2] of the MORTPAK4 produces many fluctuations after the second derivative of the death distribution function (g(x)) in many countries and because of that it was not used. Instead a new method was developed on the basis of equation (1). In fact this equation describes a non-linear regression procedure for the life table death distribution by age and the only requirement it has is to estimate the unknown parameters k, l, c, b. That was done in an Excel sheet with the aid of Excel solver in order for the sum of square errors of the fitting process to be minimized. After the estimation of the unknown parameters equation (1) was used to expand the life table death distribution for every year of life until the age

[2] http://www.un.org/en/development/desa/population/publications/pdf/mortality/mortpak_manual.pdf

of 110 years. Afterwards the First Exit Time Theory was applied in the same Excel file with the application of formulas (2) and (3)

3 Results

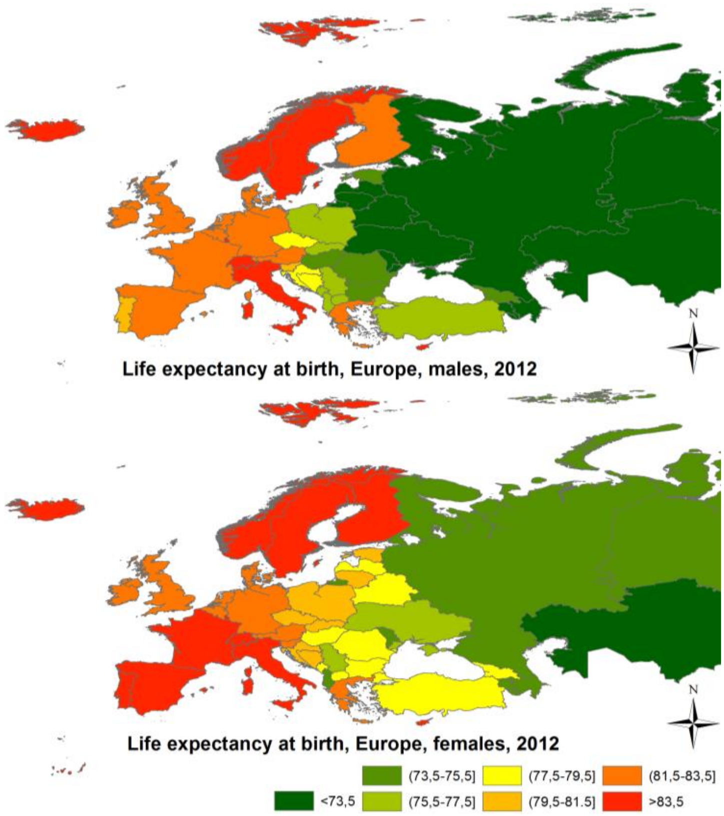


Fig. 2. Life expectancy at birth (WHO), European countries, 2012.

Based on the WHO life table results, it is quite obvious that in both genders, the observed variability of life expectancy at birth (LEB) is high among the countries of Europe (Figure 2, APPENDIX 1). In general and with few exceptions, LEB tends to increase from the eastern parts to the western and northern parts of the continent. Also high is LEB in Greece and Cyprus. These spatial trends reflect the different levels of economic and social development and are related to the socio-economic and political history of Europe, especially in the former socialistic countries, where massive systemic transformations were observed after the 1990s.

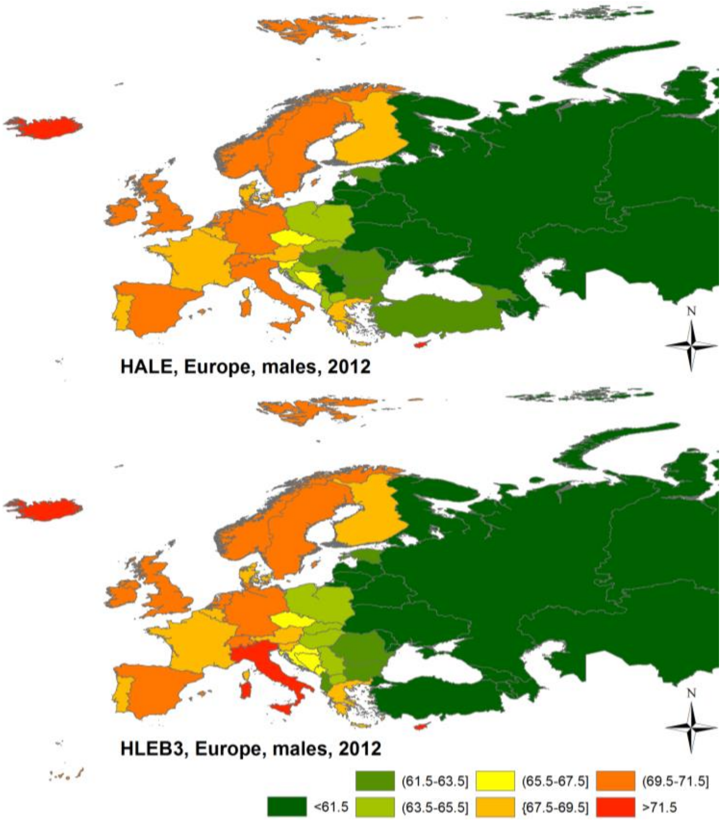


Fig. 3. HALE (WHO) and HLEB3, males. 2012.

A similar and quite variable picture emerges concerning males' Healthy Life Expectancy (HALE WHO) among the European countries in 2012 (Figure 3). The eastern parts of the continent are clearly distinguishable. As happens with the LEB, an intermediate but rather variable zone is formed by the ex-socialist countries of South-Eastern Europe all the way up to Poland and by Turkey. In the western parts of this zone HALE tends to become higher. The best HALE values are found in the rest of Europe, though variability is also high there. Clearly then, a health division exists between Eastern and the rest of Europe, including Greece and Cyprus, and this is in accordance with previous findings, in which people living in Eastern Europe self-report the worst health in the whole continent [12]. A similar picture emerges from the HLEB3 distribution; however in some countries differences are observed between the two measurements. Some of these differences are due to the way that the HALE and HLEB3 boundaries were set in order for the maps to be created and thus they are not significant, as for example happens in Italy where the difference between the rounded value of HALE[3] and HLEB3 was only +0.6 years. The differences found among the two measurements in the European countries are seen in Figure 4.

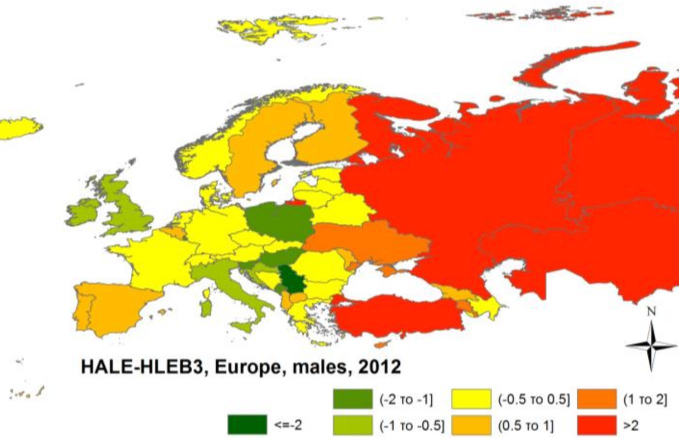


Fig. 4. HALE (WHO) and HLEB3 differences, males. 2012.

In 36 out of 48 European countries (for Liechtenstein, the Vatican City and the area of Kosovo there are no data available in the WHO database) these differences are small: ±1 year. In those countries Poland should be added (difference of -1.01 year). The confidence intervals of the WHO estimations would be very useful information in order to compare the two estimates. Unfortunately, such information is not available except for the HALE figures published by Salomon et al. [14] for the year 2010, where on average the confidence intervals ranged by ±2 years from the reported values. Despite being high, if similar confidence intervals hold for year 2012, obviously the HALE and HLEB3 concur in the very great majority of European countries: additionally, in 8 of them the differences found were between ±(1-2] years. These countries are either small (Cyprus, Luxembourg and Malta) or of the former Eastern Europe (Slovenia, Montenegro, Hungary, Armenia and Ukraine and Poland). The greater differences are found in Kazakhstan (+2.6), Russia and Turkey (+3.1).

A glimpse of this situation is given in Figure 5, where the linear relationship between HALE and HLEB3 is obvious. However, because of the deviances described above the coefficient of determination R^2 of the linear equation is 94%, which in fact is already rather high.

[3] WHO does not publish HALE statistics with decimal precision and all the estimates are rounded to the nearest integer unit.

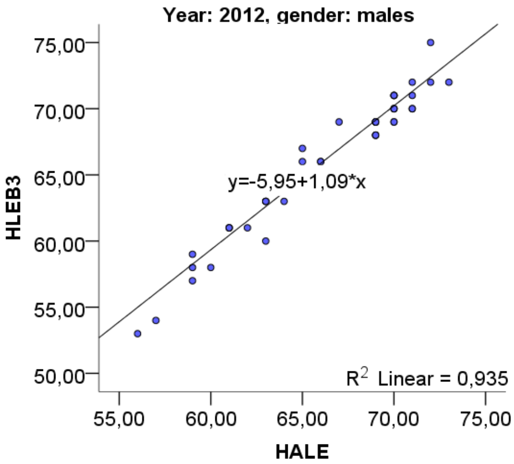


Fig. 5. HLEB and HALE scatter plot, males. 2012.

In females the confidence intervals given by Salomon et al. [14] are higher than males: on average ±2.3 years from the reported HALE. So the uncertainty is higher in females, which, besides the differential patterns of mortality and morbidity among the two genders, possibly reflects the different perceptions about disability and disease among the two genders. In the European Social Survey of 2003 it was found that men rated their health better than women in 20 of the 21 countries studied (the one being Finland) and the differences among the two genders were significant in 13 countries [12]. However, Frederiksen et al. [5], who used the grip strength as a predictor of disability, morbidity and mortality in the two sexes, found that the mean grip-strength of 80-year-old men corresponds to the mean grip-strength of a 45-year-old woman. It is about what Christensen [3] called "male-female health survival paradox". Generally men are stronger, report fewer diseases and have fewer limitations in the daily activities at older ages while women, in terms of mortality, are healthier than men. Several biological, social and psychological interpretations have been proposed for this contradiction. Not only that, but also many of the differences which are found in morbidity between the two sexes, concern variations in the definitions, diagnostic procedures and age related changes in incidence rate for many diseases (like for example coronary heart disease). Furthermore, the severity of diseases may also interfere with male-female differences. Women in general have diseases of lower risk, like migraine, arthritis and other musculoskeletal and autoimmune diseases, while men have an earlier and higher incidence of cardiovascular diseases [3].

In any case, Salomon et al. [14] state that the uncertainty intervals of their estimations depend on their knowledge of mortality rates which in some countries is uncertain. That leads to elevated uncertainty in both HALE and LEB estimations. For the females living in European countries the uncertainty could be as high as ±3 years in Iceland and as low as about ±2 years in Ukraine and Russia. On a global level, another limitation of their method is related to the use of sibling history data for 57 countries, including 25 in which such data were the only source of information on adult mortality. As for the estimation of the non-fatal outcomes, the crucial point in their analysis was how to combine information about the prevalence of 1160 disease and injury sequelae into a single summary measure which depends on the validity of the disability weights used. These data were based on 30,000 respondents from population surveys and a web-based survey (see [13]). They also stated that the findings for health severity across a very diverse sample were quite consistent but at the same time some variability remains and the differences in disability weights could be sensitive to such information. However, they focused their efforts to the lay descriptions for the majority of conditions, as they omitted some aspects of health states in the interest of simplicity, comprehensibility and feasibility. Then the effort was to ensure consistency in language across conditions and to avoid ambiguous terms. However, there is much debate on whether or not the self-reported severity of disease is independent from the socio-economic and cultural environment. For example, Allotey et al. [1] found that the severity of paraplegia, beyond its clinical manifestation is affected by several contextual factors like culture, the socio-economic status and gender (see also [22], [8]). In the European Social Survey it was found that the older persons report worse health than younger ones. Persons saying that they live comfortably on present income, report better health than those who find it difficult to live on present income [12]. Even more, it was found that the type of welfare state regime appeared to account for approximately half of the national-level variation of health inequalities between the European countries. People with Scandinavian and Anglo-Saxon welfare regimes were observed to have better self-perceived general health in comparison to East European welfare regimes [4].

Despite that, Salomon et al. (2012) emphasized the consistency of the results among different cultural environments and thus the disability weights estimated by the Global Burden of Disease Study 2010 [6] were considered to be universal. In that way it was not addressed that several complications may arise because of the use of web and household surveys. Self-selection and under coverage problems [2] are common in web surveys. Even more, the small sample size (30,000 persons in total) could give biased estimates of the prevalence of disease and sequelae as the variability of the cultures, of the socio-economic environments, the educational levels and the differential access to the health system of the individuals is huge at global and continental or even state level.

In any case the picture emerging from HALE WHO estimations and the HLEB3 from the First Exit Time Theory concerning the female population of the European continent are similar (Figure 6, APPENDIX 1).

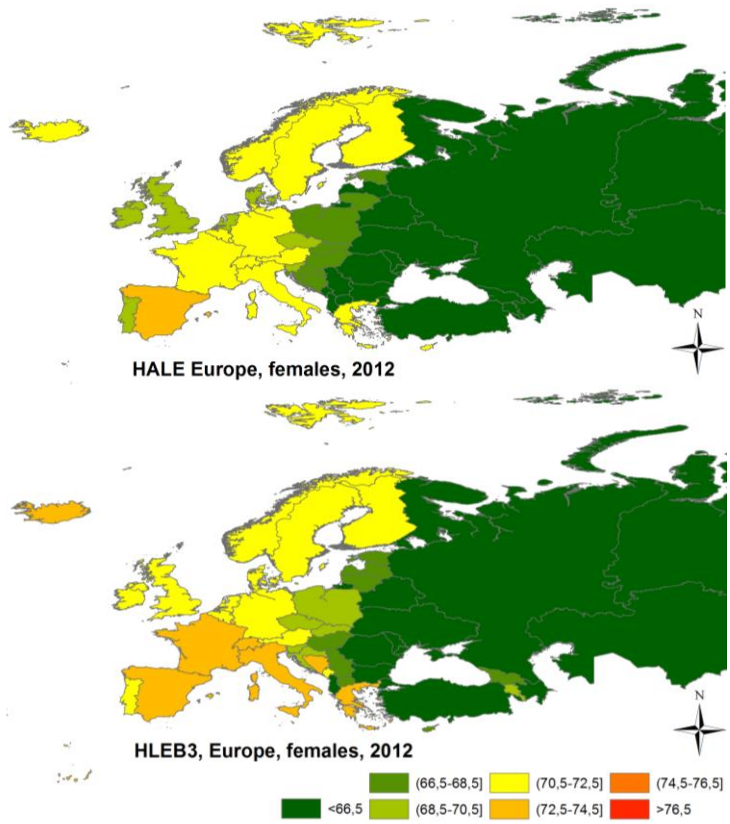


Fig. 6. HALE (WHO) and HLEB3, females. 2012.

The eastern-western division of the health of the European populations still holds. However, the differences found in the two maps of Figure 6 are greater than the analogous for males. These differences are summed in Figure 7, where it is obvious that in 16 countries the differences between HALE and HLEB3 are in the range of ±1year and in 18 in the range of ±(1-2] years. In one (Kazakhstan is 2.43 years) and in 11 is more than –2 years. If the uncertainty intervals of WHO estimations for 2012 are analogous with those reported by Salomon et al. (2012b) for 2010, then deviations outside these intervals in the two measurements reported here are found only in Greece, Portugal, Luxembourg, Belarus, Croatia and Azerbaijan. The higher deviations, which are found in the female population of the countries studied are seen in the scatter diagram in Figure 7, where despite the fact that the relationship between HALE and HLEB3 is linear and strong (R^2=0,88) it is somewhat weaker than that found for males.

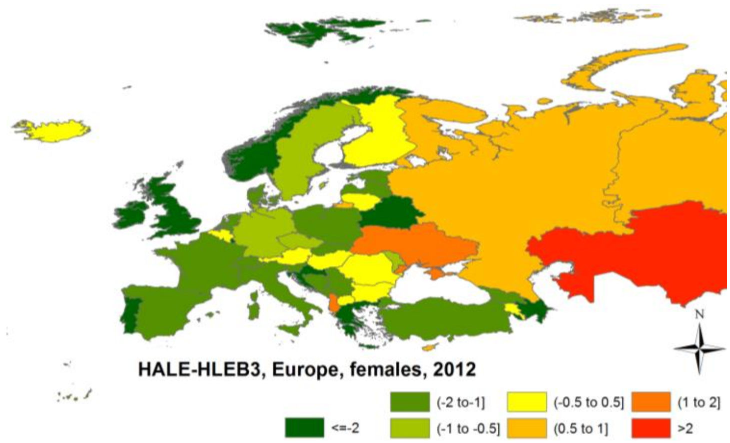


Fig. 7. HALE (WHO) and HLEB3 differences, females. 2012.

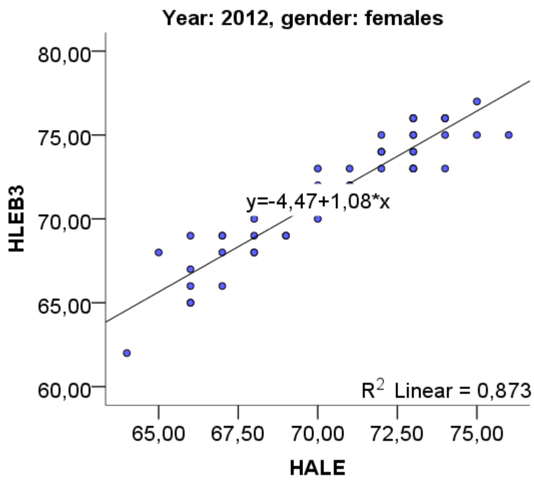


Fig. 7. HLEB and HALER scatter plot, females. 2012.

4 Conclusions

The World Health Organization estimations of Healthy Life Expectancy and the HLEB3 levels according to the First Exit Time Theory were compared based on life table data for the year 2012 published by WHO. Results indicate that in general the two methods are in accordance to each other. However, the rather complex methodology and the data requirements of the WHO method increase the difficulties of its implementation in different countries, regions and time points. In general it is a time consuming methodology, which is based on variety of data sources for estimating mortality and morbidity; however these data sources would need to be evaluated and especially the survey methods which may lead to biased estimations of healthy life expectancy.

On the other hand, the only requirement of the First Exit Time Theory is related to the availability and quality of mortality rates. If these data are available then the healthy life expectancy of any population can be immediately calculated. This gives the opportunity to policy makers, public organizations and several other institutions to have a clear picture of the current situation in order to develop their strategies and interventions. Even more, the first exit time theory can serve positively in the understanding of both mortality and morbidity in the past and explain some crucial elements of the demographic and health transition.

However, it has to be noted that the discrepancies between the two methods were higher in the female population than in the male and this may indicate the need for the further development of the first exit time theory. However, these differences can also be the result of the differential perceptions about health found among the two genders, as discussed previously in the text and in that way not significant. A further direction in the development of the First Exit Time Theory will be the evaluation of the term λ of the equation (2) in different populations and genders, as well as its closer connection with the epidemiological characteristics of a population and especially the causes of death and its health transition state.

References

1. P. Allotey, D. Reidpath, A. Kouame and R. Cummins. The DALY, context and the determinants of the severity of disease: an exploratory comparison of paraplegia in Australia and Cameroon. *Social Science & Medicine,* **57,** 949–958, 2003.
2. J. Bethlehem. Can we make official statistics with self-selection web surveys? Proceedings of Statistics Canada Symposium 2008. Data Collection: Challenges, Achievements and New Directions Statistics Canada. 2008.
3. K. Christensen. Human biodemography: some challenges and possibilities for aging research. *Demographic Research*, 19, 43, 1575-1586, 2008.
4. T. A. Eikemo, C. Bambra, K. Judge and K. Ringdal. Welfare state regimes and differences in self perceived health in Europe. *Social Science & Medicine,* 66, 2281-2295, 2008.
5. H. Frederiksen, H., J. Hjelmborg, J. Mortensen, M. Mcgue, J. W. Vaupel and K. Christensen. Age trajectories of grip strength: Cross-sectional and longitudinal data among 8,342 Danes aged 46 to 102. *Annals of Epidemiology,* 16, 7, 554–562, 2006.
6. Global Burden of Disease Study 2010. Global Burden of Disease Study 2010 (GBD 2010) Disability Weights. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2012.
7. J. Janssen and C. H. Skiadas. Dynamic modelling of life-table data, *Applied Stochastic Models and Data Analysis*, 11, 1, 35-49, 1995.
8. M. Leonardi, J. Bickenbach, T. B. Ustun, N. Kostanjsek and S. Chatterji, on

behalf of the MHADIE Consortium. The definition of disability: what is in a name? *Lancet*; 368: 1219–21, 2006.

9. C. D. Mathers, J. Salomon, C. J. L. Murray and A. Lopez. Alternative summary measures of average population health. In *Health systems performance assessment: debates, methods and empiricism*. (Murray C.J.L and Evans D., eds.). Geneva: World Health Organization; 2003
10. C. J. L. Murray, M. Ezzati et al. GBD 2010: a multi-investigator collaboration for global comparative descriptive epidemiology. Lancet, 380, 2055-2058, 2012.
11. C. J. L. Murray, M. Ezzati et al. GBD 2010: design, definitions, and metrics. *Lancet*; 380, 2063–2066, 2012b.
12. K. M. Olsen and S. A. Dahl. Health differences between European countries. *Social Science & Medicine,* 64, 1665–1678, 2007.
13. J. A. Salomon, T. Vos et al. Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *Lancet*; 380, 2129–2143, 2012.
14. J. A. Salomon, H. Wang et al. Healthy life expectancy for 187 countries, 1990-2010: a systematic analysis for the Global Burden Disease Study. *Lancet*; 380, 2144–2162, 2012b.
15. C. H. Skiadas. Life expectancy at birth and forecasts in the Netherlands (females). In *The Health State function of a population.* (Skiadas, C. H. and Skiadas, C., eds). 1st ed. Athens, 47-67, 2012a.
16. C. H. Skiadas. The Health State Function, the Force of Mortality and other characteristics resulting from the First Exit Time Theory applied to Life Table Data. In *The Health State function of a population.* (Skiadas, C. H. and Skiadas, C., eds). 1st ed. Athens, 69-92, 2012b.
17. C. H. Skiadas and C. Skiadas. Estimating the Healthy Life Expectancy from the Health State Function of a Population in Connection to the Life Expectancy at Birth. In *The Health State function of a population.* (Skiadas, C. H. and Skiadas, C., eds). 1st ed. Athens, 97-109, 2012.
18. C. H. Skiadas and C. Skiadas. *Demographic and Health indicators for 193 countries of the World Health Organization and the United Nations. Second supplement of the book The Health State Function of a Population.* Athens, 2014a.
19. C. H. Skiadas and C. Skiadas. First time exit problem. In *International Encyclopedia of Statistical Science* (Miodrag, L. ed). Springer Verlag, Berlin Heilderberg, 521-523. 2014b.
20. C. H. Skiadas and C. Skiadas. The First Exit Time Theory applied to Life Table Data: the Health State Function of a Population and other Characteristics, *Communications in Statistics-Theory and Methods,* 34, 1585-1600, 2014c.
21. D. F. Sullivan. A single index of mortality and morbidity. HSMHA Health Reports, 86, 347-354, 1971.
22. M. Schwarzinger, M. E. Stouthard, K. Burstrom and E. Nord. Cross-national agreement on disability weights: the European Disability Weights Project. *Population Health Metrics,* 1, 9, 2003.
23. T. M. Vos, A. D. Flaxman et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*; 380: 2163–2196, 2012.
24. WHO. Department of Health Statistics and Information system. WHO methods and data sources for the global burden of disease estimates 2000-2011. Global

Health Estimates Technical Paper WHO/HIS/HSI/GHE/2013.4. November, 2013
25. WHO. WHO methods for life expectancy and healthy life expectancy. Global Health Estimates Technical Paper WHO/HIS/HSI/GHE/2014.5. March, 2014

APPENDIX 1: LEB, HALE and HLEB3 by gender.

	Males							
	2012			2010 (data source: [14]				
country	LEB	HLEB	HLEB3	LEB	Uncertainty intervals	HALE	Uncertainty intervals
Cyprus	79.6	73	71.6	77.6	(77.1-78.1)	67.1	(64.7-69.1)
Iceland	81.3	72	72.0	80	(79.4-80.6)	66.9	(63.7-69.6)
Italy	80.2	71	71.6	78.9	(78.7-79.1)	68.3	(66.1-70.2)
Switzerland	80.7	71	71.4	79.7	(79.5-79.8)	69.1	(67.0-71.1)
Spain	79.2	71	70.5	78.4	(78.2-78.7)	68.8	(66.9-70.5)
Sweden	80.0	71	70.3	79.2	(79.0-79.4)	68	(65.8-70.1)
Luxembourg	79.7	70	71.2	78	(77.5-78.6)	66.9	(64.3-69.1)
Malta	78.9	70	71.1	77.1	(76.6-77.6)	66.7	(64.4-68.6)
Ireland	78.9	70	71.0	77.6	(77.4-77.9)	67.2	(64.9-69.3)
UK	78.8	70	70.5	77.8	(77.8-77.9)	67.1	(64.8-69.0)
Norway	79.5	70	69.9	78.5	(78.3-78.7)	66.3	(63.7-68.5)
Netherlands	79.1	70	69.7	78.5	(78.4-78.6)	67.9	(65.8-69.7)
Germany	78.4	70	69.6	77.5	(77.3-77.7)	67.1	(65.0-69.0)
Andorra	79.3	70	69.1	79.8	(78.8-81.0)	68.3	(65.6-70.6)
Denmark	77.7	69	69.4	76.8	(76.6-77.0)	66.3	(64.1-68.2)
France	78.6	69	69.3	77.5	(77.2-77.8)	67	(64.9-68.9)
Austria	78.3	69	69.2	77.7	(77.5-77.9)	67	(64.7-69.0)
Greece	77.8	69	69.0	77.1	(76.8-77.4)	67	(64.7-68.8)
Finland	77.7	69	68.3	76.8	(76.6-77.0)	65	(62.5-67.2)
Portugal	77.4	69	68.2	76.3	(76.2-76.5)	66.4	(64.0-68.2)
Belgium	77.8	69	68.2	76.7	(76.4-77.1)	66.5	(64.4-68.4)
Slovenia	77.0	67	68.6	75.9	(75.6-76.2)	65.7	(63.6-67.6)
Czech Republic	75.0	66	66.2	74.3	(74.2-74.5)	64.8	(62.9-66.5)
Bosnia and Herzegovina	74.6	66	66.2	74.1	(73.9-74.4)	64.4	(62.3-66.2)
Montenegro	73.5	65	66.6	73	(72.2-73.6)	63.3	(61.1-65.3)
Croatia	74.1	65	65.9	73.4	(73.2-73.6)	63.6	(61.5-65.4)
FYROM	73.5	65	64.5	72.8	(72.5-73.0)	63.2	(61.0-65.2)

Poland	72.8	64	65.0	72.1	(72.0-72.2)	62.8	(60.7-64.4)
Slovakia	72.3	64	64.2	71.6	(71.4-71.7)	62.4	(60.5-64.0)
Albania	72.6	64	63.2	72	(69.2-74.9)	62.5	(59.7-65.3)
Hungary	71.3	63	64.4	70.4	(70.3-70.6)	61.1	(59.2-62.9)
Estonia	71.3	63	63.4	70.6	(70.3-71.0)	61.7	(59.8-63.4)
Romania	70.5	63	63.2	70.1	(70.0-70.2)	61.4	(59.6-63.1)
Bulgaria	70.8	63	62.7	70.1	(69.9-70.3)	61.5	(59.5-63.2)
Turkey	71.7	63	59.9	71.2	(69.5-73.0)	61.8	(59.5-64.3)
Georgia	70.5	62	61.3	67.4	(66.0-68.7)	59.3	(57.2-61.2)
Serbia	72.0	61	65.2	74	(73.7-74.2)	64	(61.9-66.0)
Azerbaijan	69.4	61	61.5	68.9	(67.6-70.2)	59.9	(57.7-62.0)
Latvia	69.0	61	61.1	68.9	(68.6-69.2)	60	(58.1-61.7)
Lithuania	68.4	61	61.0	68.7	(68.5-68.9)	60	(58.0-61.7)
Armenia	67.2	60	58.4	68.9	(67.2-70.5)	59.9	(57.7-62.3)
Belarus	66.6	59	58.6	64.1	(63.4-64.9)	56.4	(54.6-58.1)
Moldova	66.5	59	58.3	65.5	(65.0-65.8)	57.5	(55.7-59.1)
Ukraine	65.7	59	57.2	64.5	(63.5-65.3)	56.6	(54.9-58.2)
Russia	63.4	57	53.9	63.1	(62.8-63.3)	55.4	(53.7-56.8)
Kazakhstan	63.2	56	53.5	61.3	(59.1-63.3)	53.9	(51.6-56.0)
				Females			
Cyprus	83.9	76	75.0	82.9	(82.4-83.4)	70.6	(67.8-73.0)
Spain	85.2	75	76.6	84.2	(84.0-84.4)	73	(70.9-74.8)
Andorra	85.7	74	76.1	85.2	(84.2-86.2)	72.2	(69.4-74.7)
France	84.9	74	75.1	84.3	(84.0-84.5)	71.9	(69.5-74.0)
Italy	85.0	74	75.9	83.9	(83.7-84.1)	71.9	(69.6-73.9)
Switzerland	85.1	74	76.0	84.5	(84.3-84.7)	72.4	(70.0-74.6)
Austria	83.4	73	73.4	83.3	(83.2-83.5)	71.2	(68.7-73.5)
Belgium	83.0	73	73.0	82.3	(81.9-82.6)	70.6	(68.4-72.7)
Finland	83.6	73	73.5	83.3	(83.1-83.6)	69.6	(66.8-72.0)
Germany	83.1	73	73.5	82.8	(82.6-83.1)	70.9	(68.7-72.9)
Greece	83.5	73	76.3	82.1	(81.9-82.4)	70.4	(68.0-72.5)
Iceland	83.6	73	73.5	84.4	(83.7-85.0)	69.9	(66.7-72.7)
Ireland	83.3	73	75.3	82.2	(81.9-82.4)	70.5	(68.2-72.7)
Luxembourg	84.1	73	75.8	82.2	(81.7-82.8)	69.9	(67.1-72.1)
Portugal	84.0	73	76.0	82.3	(82.2-82.5)	70.7	(68.3-72.9)

Slovenia	83.5	73	74.8	82.5	(82.2-82.9)	70.7	(68.3-72.9)
Sweden	83.6	73	73.5	83.5	(83.4-83.7)	71.2	(68.8-73.4)
Denmark	81.9	72	73.8	81	(80.8-81.3)	69.5	(67.1-71.6)
Malta	82.8	72	74.5	83	(82.4-83.6)	70.6	(67.8-73.1)
Netherlands	82.9	72	73.3	82.6	(82.4-82.7)	70.2	(68.0-72.2)
Norway	83.6	72	74.3	83.1	(82.9-83.4)	69.7	(67.1-72.0)
UK	82.8	72	74.3	81.9	(81.8-82.0)	70.1	(67.8-72.1)
Czech Republic	81.0	71	71.7	80.7	(80.5-80.8)	69.6	(67.3-71.6)
Estonia	81.4	71	72.4	80.6	(80.2-81.0)	69.3	(67.0-71.2)
Poland	81.0	71	72.5	80.5	(80.4-80.6)	69.3	(66.7-71.5)
Bosnia and Herzegovina	79.6	70	71.7	78.8	(78.5-79.0)	68.1	(65.8-70.2)
Croatia	81.0	70	73.1	79.9	(79.7-80.1)	68.3	(65.9-70.3)
Lithuania	79.6	70	70.1	79.3	(79.1-79.6)	68.4	(66.1-70.4)
Slovakia	79.9	70	71.2	79.1	(78.9-79.3)	68.3	(66.1-70.3)
Hungary	78.7	69	69.4	78.4	(78.2-78.5)	67.3	(65.0-69.2)
Romania	78.1	69	69.4	77.6	(77.4-77.7)	67.3	(65.1-69.2)
Bulgaria	77.7	68	67.8	77	(76.8-77.2)	66.8	(64.7-68.7)
FYROM	77.6	68	68.3	77.2	(77.0-77.5)	66.4	(64.0-68.4)
Georgia	77.8	68	69.9	77.9	(76.9-78.9)	66.9	(64.5-69.1)
Latvia	78.6	68	69.4	78.5	(78.2-78.7)	67.2	(64.8-69.3)
Montenegro	78.3	67	68.5	78.2	(77.5-78.9)	66.1	(63.4-68.4)
Serbia	77.2	67	68.7	79.5	(79.2-79.8)	68	(65.5-70.2)
Turkey	78.4	67	68.9	77.7	(75.9-79.3)	66	(63.5-68.4)
Ukraine	75.6	67	65.9	74.9	(74.2-75.4)	64.9	(62.8-66.7)
Albania	75.3	66	64.8	78.1	(75.9-80.2)	67	(64.4-69.7)
Armenia	75.1	66	66.3	78.5	(77.4-79.6)	67.2	(64.7-69.5)
Belarus	77.8	66	69.1	76	(75.5-76.5)	65.6	(63.3-67.6)
Moldova	74.6	66	66.7	74.6	(74.2-74.9)	64.6	(62.4-66.7)
Russia	75.1	66	65.2	74.7	(74.4-74.9)	64.5	(62.5-66.2)
Azerbaijan	75.1	65	68.0	76.2	(74.9-77.4)	65.1	(62.6-67.3)
Kazakhstan	72.4	64	61.6	72.2	(70.6-73.7)	62.4	(59.9-64.6)

Modeling of some physiological parameters and feeding behavior of sheep in response to heat stress: A meta analytic approach

Imen Belhadj Slimen[1,2] and Taha Najar[1,2]

[1] Department Department of Animal, Food and Halieutic Resources, National Agronomic Institute of Tunisia, 43 Av Charles Nicolle 1082 Tunis, Tunisia
(E-mail: belhadj_slimen_imen@yahoo.fr, najar.taha@inat.agrinet.tn)
[2] Laboratory of Materials, Molecules and Application, Preparatory Institute for Scientific and Technical Studies, road of Sidi Bou Said, B.P:51, 2075, La Marsa, Tunis, Tunisia

Abstract. This study aims to evaluate the effect of environmental-induced heat stress on some physiological parameters and feeding behavior of sheep. The results of published papers related to the topic of heat stress effects on sheep were used to generate a database. Studied parameters were evaluated using correlation, PCA and covariance analysis. Simple and multiple regressions were carried out to quantify the effect of heat stress on the studied responses. Although rectal temperature (RT), respiration rate (RR) and water intake (WI) are highly and positively correlated with ambient temperature and thermo hygrometric index (THI), a negative correlation coefficient was observed for feed intake. Moreover, heat stress effect is breed dependant: temperate and crossbreeds are more affected by heat stress than African and West-South-Asian breeds. Indeed, males are more resistant to heat load than females and castrated sheep. Increased THI values induce the increase of the RT.
Keywords: Heat stress, sheep, meta-analysis

1 Introduction

Environmental-induced heat stress jeopardizes animal welfare and compromises livestock production performances (Najar et *al*. [1]; Marai et *al*., [2]; Abdel–Hafez [3]). In sheep, heat stress was shown to increase the rectal temperature and the respiration rate of exposed animals, and to affect negatively their growth performances of lambs. Decreased body weight, average daily gain and growth rate were reported in lambs (Marai et *al*. [2]; Padua et *al*. [4]). Decreased production performances were usually thought to originate from the decreased feed intake, observed in Croix, Karakul, Rambouillet, Suffolk and Comisana sheep following heat exposure (Nardon et *al*. [5]; Padua et *al*. [4]). However, these studies were carried out using a limited number of environmental, physiological and nutritional parameters. Moreover, the huge difference in ovine breeds defines various responses to heat stress. Indeed, the sex of the animal interferes with the heat response. Therefore, the meta-analysis is an efficient tool to overcome the heterogeneity of these parameters and to make the interactions of breed and sex with the heat response more comprehensible.

16rd ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

2 Materials and methods

Data collection

A literature review was conducted to evaluate the effect of the ambient temperature and the relative humidity on ovine rectal temperature, respiration rate as well as feed and water intake. Only papers reporting THI or/and experimental temperatures and relative humidity values were considered in our study. The Table 1 records the number of papers considered for each studied parameter. Ovine breeds were classified into 4 groups: temperate breeds, African breeds, west and South African breeds and crossbreds, as detailed in Table 2. The sex of the animals was also considered: sheep were grouped as females, males and castrated males.

Parameter	n	Mean	Minimum	Maximum
THI	190	75.05	47.44	111.89
Relative humidity (%)	173	58.08	5.29	93
Temperature (°C)	173	27.21	8.96	50
Rectal temperature (°C)	191	39.18	37.07	40.86
Respiratory rate (breath/mn)	147	83.36	21.5	272
DMI (Kg/d)	48	1.53	0.5	3.1
Water intake (l/d)	16	5.33	2.3	9.3

Table 1. Statistical description of the environmental, physiological and dietary retained parameters

Ovine breeds	N	Origin
Temperate breeds	72	Spain, Sardinia, Italy, Angleterre, Brazil, Carribbean, USA, Australia, Europe
West and South Asian breeds	53	Arabie Saoudite, Jordan, Iraq, Syria, Libanon, Oman, Kuwait, Israel
African breeds	26	South Africa, Sudan, Egypt
Cross breeds	39	Cross between temperate breeds

Table 2. Description of the number and the origin of ovine breeds used in this study

Calculations and statistical analysis

THI was computed for papers reporting experimental temperatures and relative humidity values as described by Thom [6]:

$$THI = 0.8° \, TC + ((RH/100)*(TC - 14.3)) + 46.4$$

Where TC is the ambient temperature (°C) and RH is the relative humidity (%).
Correlations and Principle Component Analysis (PCA) were used to identify the parameters that affect the animal response to heat stress. The effect of heat stress was quantified using simple and multiple regressions. Covariance analysis was used to study the effect of breed and sex on the response of animal under different ambient temperatures and relative humidity values.

3 Results and discussion

Rectal temperature and respiration rate

As shown in Table 3, a significant correlation is observed between rectal temperature and the environmental parameters (temperature, relative humidity and THI) as well as breed and sex. Similarly, this observation is also available for respiration rate. The negative correlation coefficient noted between respiration rate and RH confirms studies reporting that high humidity levels reduce respiration frequency, decrease evapotranspiration and aggravate heat stress (Marai et *al*., [7]; Lin et *al*. [8]).

	Breed	Sex	Temperature	Relative humidity	THI
Rectal temperature	-0.374 ***	-0.120 .	0.543 ***	0.148 *	0.609 ***
Respiration rate	-0.171 *	-0.255 **	0.552 ***	-0.262 **	0,495 ***

Table 3. Correlation coefficients of rectal temperature and respiration rate

PCA allowed to retain CP1 and CP2 axis, which explain 62% of the variability of studied data, as shown in Table 4.

	Eigen values	Proportion of variance (%)	Cumulative variance (%)
CP1	2.98	42.6	42.6
CP2	1.35	19,4	62
Explained variance (%)		62	

Table 4. Description of the variance of CP1 and CP2 retained by PCA

ANOVA analysis of the effects of breed and sex on rectal temperature and respiration rate showed significant relations ($p< 0.0001$), as shown in Table 5. But their interaction is not statistically significant ($p<0.1$). Breed and sex explain about 23% of RT and RR variance.

	Rectal temperature (°C)	Respiration rate (breath/mn)
R^2	0.232	0.225
Model	< 0.0001	< 0.0001

Breed	< 0.0001	< 0.0001
Sex	0.052	0.005
Breed*Sex	0.645	0.978

Table 5. p-values of the parameters used in ANOVA models

Based on the results of correlation ANOVA, and PCA, a simple linear regression was established between the RT and the THI values (R^2 = 37.1%, p<0.0001). As represented in Figure 1, heat stress effect on rectal temperature is observed when THI reaches 75. If THI increases by 5 points, the RT increases by 0.19°C. Figures 2 and 3 represent the simple linear regression between ovine rectal temperatures and the THI values fitted by sex.

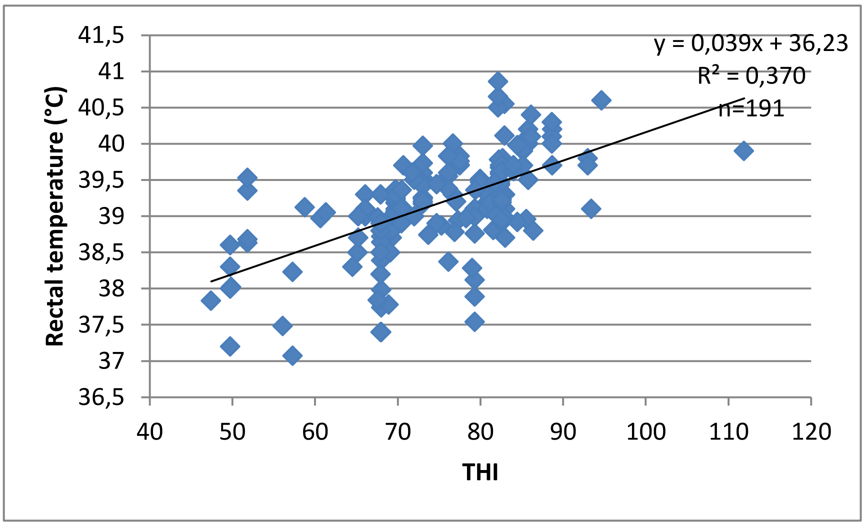


Fig 1. Plot of the rectal temperature values as a function of the THI

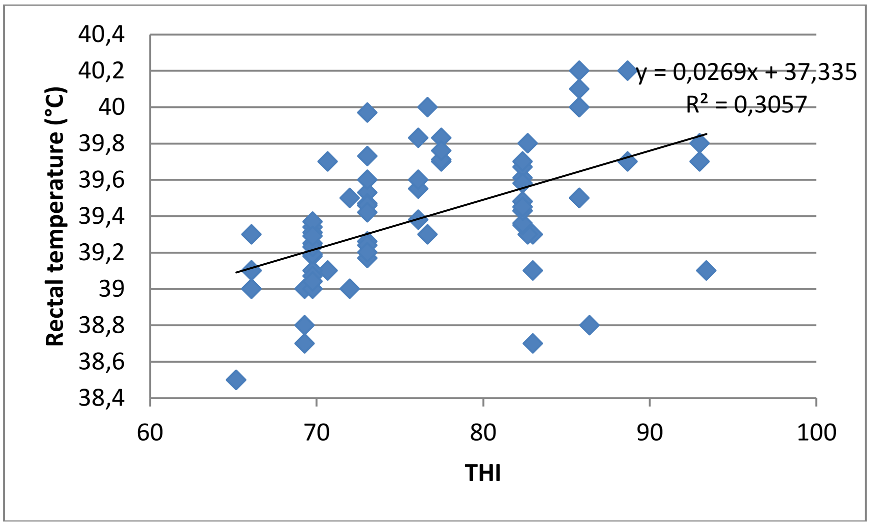


Fig 2. Plot of the rectal temperature of ewes as a function of the THI

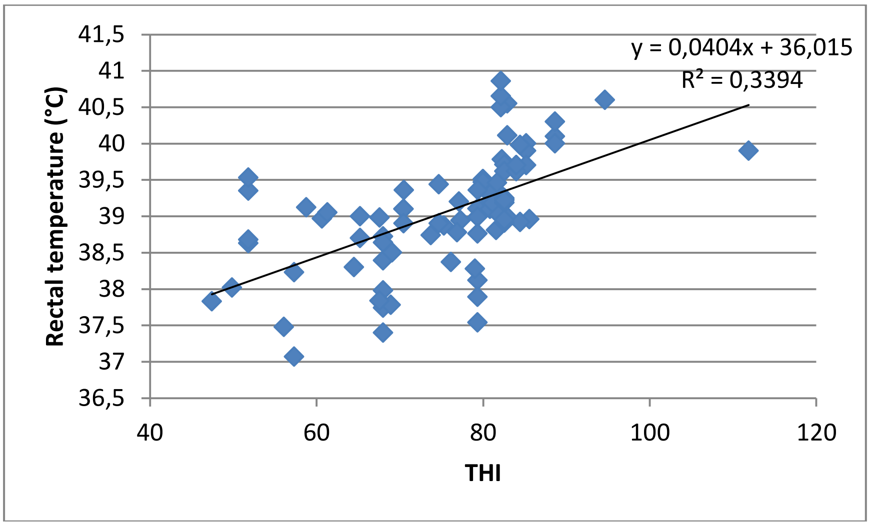


Fig 3. Plot of the rectal temperature of male ovins as a function of the THI

The RR was expressed using the ambient temperature and the relative humidity values.

RR = - 20,6 + 4,01 TC° - 0,082 RH (R^2 = 30.5%, $p<0.000$).

Linear regression results show that RT and RR increase in response to an elevation of the THI and the ambient temperature. RR values followed the same pattern with that of RT with higher RR values recorded in black animals. The

animals panted in order to increase body cooling by respiratory evaporation since the major evaporatory heat loss mechanism is panting. However, about 65% of the variability of RT and RR rate values cannot be explained based on these environmental parameters.
ANCOVA analysis showed that both breed and sex (used as covariables) affect significantly the rectal temperatures and respiration rates of the studied animals, in response to a heat load, as shown in Table 6.

	Rectal temperature (°C)	Respiration rate (breath/mn)
Model	<0.0001	<0.0001
R^2	0.515	0.507
THI	<0.0001	-
Ambient temperature	-	<0.0001
Relative humidity	-	0.082
Breed	<0.0001	<0.0001
Sex	0.672	0.000

Table 6. p-values of the parameters used in ANCOVA models

Based on the results of ANCOVA analysis, cross and temperate breeds are more susceptible to heat stress than Afrian and West - South Asian ones, as shown in Table 7. Our results are in accordance with Marai et *al*. [9] who reported that Egyptian breeds are more tolerant to heat stress than temperate ones. In Drosophila, the most heat tolerant population comes from a climate exhibiting the warmest annual mean temperature in Argentina (Fallis et *al*. [10]). Indeed, the northern populations of springtail Orchesella cincta collected from Denmark to southern Italy had a lower heat shock resistance than the southern ones (Bahrndorff et *al*. [11]). These observations provide evidence for geographical variation in heat response which is, most likely, due to a previous local phenotypic and genetic adaptation to constant, high thermal environments.

	Rectal temperature (°C)	Respiration rate (breath/mn)
Cross breeds	39.51 ± 0.09^a	90.17 ± 8.06^a
Temperate breeds	39.30^{ab}	102.34^a
West and South Asian breeds	39.09 ± 0.08^b	62.54 ± 10.21^b
African breeds	38.54 ± 0.12^c	58.87 ± 15.47^b

Table 7. Variation of the rectal temperature and the respiration rate averages as function of the geographic location of the ovine breed

Moreover, females and castrated males appear to be more affected by heat stress than males, as shown in Table 8. The highest rectal temperature values were recorded in castrated males and females. Similarly, the respiration rate of males is significantly lower (p = 0.000) than that observed in females and castrated males (respectively 73.93 ± 9.33 versus 74.23 ± 10.08 and 96.82 breath / mn).

	Number of observations	Rectal temperature (°C)
Castrated males	17	39.58± 0.12[a]
Females	82	39.38[a]
Males	91	38.93 ± 0.08[b]

Table 8. Effect of the sex on the rectal temperature response

These results are in accordance with Fadare et *al*. [12], who reported that the average rectal temperature of females was higher than males (38.83 ± 0.02°C versus 38.69 ± 0.02°C). Indeed, Sejian et *al*. [13] reported increase in RT and RR as a result of heat stress in female sheep. The pronounced differences between males and females in their responses to environmental-induced heat stress are probably linked to the hypothalamo-pituitary adrenal axis, and it has been proposed that sex differences in response to stress are due principally to the influence of sex steroids (Handa et *al*. [14]; Tilbrook et *al*. [15]).
ANCOVA analysis (THI, ambient temperature, RH, breed, sex) contributed to explain about 50% of the variability of rectal temperature and respiration rate values. The unexplained results may be due to effects of the different treatments used to alleviate the drastic effects of heat stress on sheep, such as sheering, offering shade, ventilation, selenium supplementation, etc.

Water and feed intake

Water intake (WI) and dry matter intake (DMI) are more correlated to ambient temperature and relative humidity than to THI. Therefore, multiple regressions were used to modelize the relationship between these variables.

WI (l) = 0.555 + 0.194*TC + 0.002*RH (R^2=0.450, p = 0.02, RMSE = 4.30)
DMI (Kg) = 1.168 – 0.016*TC + 0.016*RH (R^2=0.189, p = 0.01, RMSE = 0.436)

If the ambient temperature increases by 5°C, water intake increases by 0.97 l and DMI reduces by -0.08 Kg.

Water is one of the most important nutrients, and is involved in many physiological functions, such as the thermoregulation. As it was reported previously, exposure of sheep to hot environments increases their water consumption from 2 Kg water/Kg DM at temperatures between 0 and 15°C, to 3 Kg water/Kg DM at temperatures above 20°C (Conrad [16]). Elevated water consumption under hot environments is accompanied by an increased water turnover (Murad et *al*. [17]). Under heat stress, sheep use water to decrease their body temperature by the loss of heat to ingested water.
The regression coefficient between DMI and temperature is negative, indicating a depressed feed intake in response to elevated temperatures. However, this coefficient is positive between water intake and temperature, because of the

increase in water consumption under high temperatures. Heat stress up-regulates the secretion of two adipokines: leptin and adiponectin, which stimulate the hypothalamic axis and result in a reduced feed intake (Al-Azraqui [18], Morera et *al*. [19]). This form of caloric restriction allows to reducing heat generation.
The model proposed by the ANCOVA analysis of DMI (Table 8) improved the significance of the results (R^2=0.765) and revealed that DMI is affected significantly (<0.0001) by breed. Once again, temperate breeds recorded the lowest DMI (0.534 ± 0.16 Kg), whereas the DMI of West-South Asian breeds is 3 folds higher (1.586 ± 0.17 Kg). Maintaining normal feed intake values seems to be an indicator of heat tolerance.

	DMI (Kg)
Model	<0.0001
R^2	0.765
Ambient temperature	<0.0001
Relative humidity	0.001
Breed	<0.0001

Table 8. p-values of the parameters used in the DMI ANCOVA model

Conclusions

Exposure to high ambient temperature augments the effort to dissipate body heat, resulting in an increase of respiration rate, body temperature and consumption of water, and a decline in feed intake. This study showed that West-South Asian and African breeds are more tolerant to environmental-induced heat stress than temperate and cross breeds. Heat tolerance suggests least variation from normal in traits, such as rectal temperature, respiration rate and feed intake, when raised under heat stress conditions. Hence, the selection for high production performances (growth and milk yield) is not recommended for sheep reared in arid and semi arid areas.

References

1. T. Najar, M. Rejeb, M. Ben M'Rad. Modeling the effects of heat stress on some behavior and physiological parameters in cows. In: Modelling nutrient digestion and utilization in farm animals. *Applied Stochastic Models and Data Analysis*, Wageningen Academic Publishers, The Netherlands, 130-136, 2010 .
2. I.F.M. Marai, A.M. Daader, A.M. Abdel-Samee, H. Ibrahim. Winter and summer effects and their amelioration on lactating Friesian and Holstein cows maintained under Egyptian conditions. *Proceedings of International Conference on Animal, Poultry, Rabbits and Fish Production and Health*, Cairo, Egypt, 1997.
3. M.A.M. Abdel-Hafe. Studies on reproductive performance in sheep. *Ph.D. Thesis, Faculty of Agriculture, Zagazig University, Egypt,* 2002.
4. J.T. Padua, R.G. Dasilva, R.W. Bottcher, S.J. Hoff. Effect of high environmental temperature on weight gain and food intake of Suffolk lambs reared in a tropical

environment. *Proceedings of 5th International Symposium*, Bloomington, Minnesota, USA, 809–815, 1997.
5. A. Nardon, B. Ronchi, A. Valentini.Effects of solar radiation on water and food intake and weight gain in Sarda and Comisana female lambs. *Animal Husbandry in Warm Climates*, 55, 149–150, 1991.
6. E.C. Thom. The Discomfort Index. *Weatherwise* 12, 57-59, 1959.
7. I.F.M. Marai, L.B. Bahgat, T.H. Shalaby, M.A. Abdel-Hafez. Fattening performance, some behavioural traits and physiological reactions of male lambs fed concentrates mixture alone with or without natural clay, under hot summer of Egypt. *Ann. Arid Zone*, 39, 4, 449–460, 2000.
8. H. Lin, R. Du, X.H. Gu, F.C. Li, Z.Y. Zhang. A study of the plasma biochemical indices of heat stressed broilers. *Asian-Australian Journal Animal Science*, 13, 9, 1210-1218, 2000.
9. I.F.M. Marai, A.A. El Darawany, A. Fadiel, M.A.M. Abdel Hafez. Physiological traits as affected by heat stress in sheep: a review. *Small Ruminant Research*, 71, 1-12, 2007.
10. L.C. Fallis, J.J. Fanara, T.J. Morgan. Genetic variation in heat stress tolerance among south American *Drosophila* populations. *Genetica*, 139, 1331-1337, 2011.
11. S. Bahrndorff, M. Holmstrup, H. Peterson, V. Loeschcke. Geographic variation for climatic stress resistance traits in the springtail Orchesella cincta. *Journal of Insect Physiology*, 52, 9, 951-959, 2006.
12. A.O. Fadare, S.O. Peters, A. Yakubu, A.O. Sonibare, M.A. Adeleke, M.O. Ozoje, I.G. Imumorin. Physiological and haematological indices suggest superior heat tolerance of white-colored West African Dwarf sheep in the hot humid tropics. *Tropical Animal Health and Production*, 45, 1, 157-165, 2012.
13. V. Sejian, V.P. Maurya, M.K. Naqvi. Adaptability and growth of Malpura ewes subjected to thermal and nutritional stress. *Tropical Animal Health and Production*, 42, 1763-1770, 2010.
14. R.J. Handa, L.H. Burgess, J.E. Kerr, J.A. O'Keefe. Gonadal steroid hormone receptors and sex differences in the hypothalamo–pituitary–adrenal axis. *Hormones and Behavior*, 28, 464–476, 1994.
15. A.J. Tilbrook, A.I. Turner, I.J. Clarke. Effects of stress on reproduction in non-rodent mammals: the role of glucocorticoids and sex differences. *Reviews of Reproduction*, 5, 105–113, 2000.
16. J.H. Conrad. Feeding of farm animals in hot and cold environments. *Stress Physiology in Livestock*, CRC Press, U.S.A., 1985.
17. H.M. Murad, T.M. El-Bedawy, M.M. Shafie, S.M. Salem. Effect of heat stress and dietary roughage level on rumen function, nutrient utilization and water turnover by native Egyptian sheep, Egypt. *Journal of Animal Production*, 31 (Suppl.), 361–371, 1994.
18. A.A. Al-Azraqui. Pattern of leptin secretion and oxidative markers in heat-stressed-pigeons. *International Journal of Poultry Science*. 7, 12, 1174-1176, 2008.
19. P. Morera, L. Basirico, K. Hosoda, U. Bernabucci.. Chronic heat stress up-regulates leptin and adiponectin secretion and expression and improves leptin, adiponectin and insulin sensivity in mice. *Journal of molecular endocrinology*. 48, 129-138, 2012.

Occupational and educational gender segregation in Southern Europe

G. Stamatopoulou, M. Symeonaki, and C. Michalopoulou

Department of Social Policy, Panteion University of Social and Political Sciences, 136 Syggrou Av., 176 71, Athens, Greece
(E-mail: gl.stamat@gmail.com, msymeon@panteion.gr, kmichal@panteion.gr)

Abstract. The present paper focuses on the measurement of gender segregation in education and occupation, using data drawn from the 2011 European Survey on Income and Living Condition (EU-SILC). The concepts of distance and similarity are used in the methodology. We use a number of distance and similarity indices to capture the relationship between men's educational and occupational distribution and the respective distribution of women and to show whether and to what extent there is a differentiation in educational and occupational attainments of men and women. Additionally, a cross-national comparison of levels of gender segregation is provided between Southern European counties (Greece, Italy, Portugal and Spain).
Keywords: gender segregation, social distances, southern Europe, ESeC, ISCED.

1 Introduction

In recent decades, gender segregation, i.e. the dissimilarities observed in the social outcomes of individuals due to their gender, has become an issue of utmost concern in Europe. According to the official statistical reports, gender differences in education and labour market are still present, even though major reforms have undergone to promote more equal opportunities for both men and women.

The present paper focuses on the measurement of gender educational and occupational segregation, using data drawn from Eurostat, 2011 European Union Statistics on Income and Living Condition (EU-SILC). Replacing the European Panel Survey since 2003, EU-SILC provides timely and comparable cross-sectional, multidimensional and longitudinal microdata on income, poverty, social exclusion and living conditions anchored in the European Statistical System. Covering all member states of the European Union, the main goal of the survey is to document the living conditions of households and their members, as well as the social and economic characteristics that affect these conditions.

The concepts of distance and similarity are used in the methodology. In general, similarity measures and coefficients are used to describe how similar two data points (clusters, distributions, samples, etc) are, whereas distance or dissimilarity measures are used to examine how dissimilar two data points are.

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

In this respect, we use a number of distance and similarity indices to capture the relationship between men's educational and occupational distribution and the respective distribution of women and to show whether and to what extent there is a differentiation in the educational and occupational attainments of men and women (see also Symeonaki & Stamatopoulou [9] and Symeonaki *et al.* [10] for an application of distance indices to the measurement of intergenerational mobility in Greece). Additionally, a cross-national comparison of levels of gender segregation is provided among Southern European counties (Greece, Italy, Portugal and Spain).

For analytical purposes, we apply the European Socio-economic Classification (ESeC scheme) and nine occupational classes are constructed, to identify the occupational level: *1) Large employers, higher grade professionals; 2) Lower grade professionals, higher supervisory/technicians; 3) Intermediate occupations; 4) Small employers and self-employed (non-agriculture); 5) Small employers and self-employed (agriculture); 6) Lower supervisors and technicians; 7) Lower sales and service occupations; 8) Lower technical occupations; 9) Routine occupations.*

To identify the educational level, the male and female classification is based on their completed level of education, according to the latest version of International Standard Classification of Education (ISCED-11). The educational attainment has been recoded into four educational states, as indicated in Table 1. Moreover, the data was weighted by applying the design weight, as is required by probability sampling theory.

ISCED	**Educational categories**	**Description**	**States**
0-1	Less than lower secondary	Primary education (un)completed	1
2	First stage of secondary completed	3-year lower secondary education	2
3	Second stage of secondary completed	3-year upper secondary education	3
4-6	Advanced education	Post-secondary or tertiary education	4

Table 1. Educational categories according to ISCED11 classification used by EU-SILC.

The paper has been organised in the following way. Section 2 provides the preliminaries and notation employed in the paper, while Section 3 presents the distance and similarity measures, as well as the ratio index of gender occupational and educational segregation. Section 4 presents our main findings, whereas Section 5 summarises the conclusions.

2 Preliminaries and notation

The mathematical notion of distance introduced in Frechet[4] and Hausdorff[6] is given in Definition 1, while the definition of a metric distance is given by Definition 2 (see also Anderberg, 1973; Zhang & Srihari, 2003).

Definition 1. Let X be a set. A distance on X is a function $d : X \times X \to R$, where R is the set of real numbers, if and only if it satisfies the following conditions, $\forall x, y \in X$:

1. $d(x, y) \geq 0$, (non-negativity or positivity axiom),
2. $d(x, y) = 0$, if and only if $x = y$, (reflexivity axiom), and
3. $d(x, y) = d(y, x)$, (symmetry axiom).

Definition 2. Let X be a set. A metric distance on X is a function $d : X \times X \to R$, where R is the set of real numbers, if and only if it satisfies the following conditions, $\forall x, y \in X$:

1. $d(x, y) \geq 0$, (non-negativity or positivity axiom),
2. $d(x, y) = 0$, if and only if $x = y$, (identity axiom),
3. $d(x, y) = d(y, x)$, (symmetry axiom), and
4. $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

Table 2 provides a summary of the basic notation used in the remaining of the paper.

Table 2. Summary of the notation used

Notation	**Description**
d	denotes the number of classes, i.e. the occupational/educational space, denoted by $S = \{1, 2, ..., d\}$,
m_i	is the number of men in the $i - th$ class,
f_i	is the number of women in the $i - th$ class,
m	is the total number of men, i.e. $m = \sum_{i=1}^{d} m_i$,
f	is the total number of women, i.e. $f = \sum_{i=1}^{d} f_i$,
p_i^m	denotes the proportion of men in $i - th$ class, i.e. $p_i^m = \frac{m_i}{m}$,
p_i^f	denotes the proportion of women in $i - th$ class, i.e. $p_i^f = \frac{f_i}{f}$,
$\mathbf{p}_m$	denotes the vector $\mathbf{p}_m = [p_1^m, p_2^m, ..., p_d^m]$, and
$\mathbf{p}_f$	denotes the vector $\mathbf{p}_f = [p_1^f, p_2^f, ..., p_d^f]$.

3 The measurement of Gender Segregation

Following the notation introduced in Section 2 we provide a number of distance and similarity measures, which will be used in the analysis. Moreover, the ratio index of gender segregation introduced by Maria Charles and David Crusky[2] and also used by Valentova et al.[8] is computed to give some additional results.

The most widely known metric distance between two points (clusters, objects, etc) is the Euclidean distance, also called Pythagorean or ruler distance. It is the length of the line segment connecting the two points, being also the shortest distance between them. Focusing on the case of gender segregation, the distance between the vectors $\mathbf{p}_m$ and $\mathbf{p}_f$ is given by the Equation 1. The higher the values of Euclidean distance, the more dissimilar are the distributions.

$$d_{Euc} = \sqrt{\sum_{i=1}^{d} |p_i^f - p_i^m|^2}. \tag{1}$$

Manhattan distance (Equation (2)), considered by Hermann Minkowski in the 19th century, is a distance function in which the usual Euclidean distance function is replaced by a new metric, where the distance between two points is the sum of the absolute differences of their coordinates. The Manhattan metric is also known as rectilinear distance, L_1 distance or 1-norm, city block distance, Taxicab distance, Manhattan length or Fottruler distance.[7]

$$d_{Man} = \sum_{i=1}^{d} |p_i^f - p_i^m|. \tag{2}$$

Its true meaning, therefore, lies in the fact that it represents the total variation of men and women when their distribution to occupational or educational classes is considered. It is important to state here that the Manhattan distance is equal to the index introduced much later by Moir and Selby-Smith (1979), also called the MSS segregation indicator. This index was also introduced by the Organisation for Economic Cooperation and Development (OECD) in European analysis (OECD 1980, 1985), under the name WE after the OECD's report "Women in Employment". Therefore $d_{Man} = MSS = WE$.

Chebyshev distance, considered by Pafnuty Chebyshev, also called Maximum metric, sup metric or L_1 metric [7] is a metric defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension.[1]

$$d_{Cheb} = \max_i |p_i^f - p_i^m|. \tag{3}$$

Notice that, Chebychev distance considers only the part for which the difference is maximum, while Manhattan distance gives equal importance to all differences. Therefore, we could assume that Chebychev distance represents the maximum proportional variation of the occupational and educational gender-based distribution. High values of the index indicate that there is a great difference between men and women in the educational or occupational class that they differ more.

The Sorensen index (Equation 4), also known as Sorensen's similarity coefficient, is a statistical index used to compare the similarity of two samples and coincides with the Index of Dissimilarity ID, which is widely used as a measure of gender segregation, particularly in American literature.

$$d_{Sor} = \frac{\sum_{i=1}^{d} |p_i^f - p_i^m|}{\sum_{i=1}^{d} (p_i^f + p_i^m)} = \frac{\sum_{i=1}^{d} |p_i^f - p_i^m|}{2}. \tag{4}$$

Obviously, d_{Sor} equals to the Manhattan distance divided by 2 and therefore corresponds to the half of the total variation between men and women. $d_{Sor} = 0$ in case of perfect similarity, and $d_{Sor} = 1$ in case of perfect dissimilarity.

Gower's similarity index is given by Equation (5). The higher the values, the greater the dissimilarity.

$$d_{Gow} = \frac{\sum_{i=1}^{d} |p_i^f - p_i^m|}{d}. \tag{5}$$

The index of gender segregation is computed, using the Equation (6), as a supplementary tool, in order to capture the "net" sex ratio in particular occupational classes or educational levels. As mentioned in Valentova et al.[8], the index is based on the classification of both general (for a state) and specific (for every occupational or educational category) gender segregation score, while $exp(R)$ shows the extent to which women are disproportionately represented. The higher the value of the index, the higher the gender segregation.

$$R = [1/d] \sum_{i=1}^{d} \left\{ ln(f_i/m_1) - [1/n] \sum_{i=1}^{d} ln(f_i/m_i) \right\}^2. \tag{6}$$

Finally, R_i gives the ratio index score of gender segregation in particular occupational and educational levels and is computed by using Equation (7). Negative values of the index show male overrepresentation.

$$R_i = ln(f_i/m_1) - [1/n] \sum_{i=1}^{d} ln(f_i/m_i). \tag{7}$$

4 Distance and similarity measures of gender segregation

Having presented the measures, we now proceed with the presentation of the results of our analysis. Similarity and distance indices between men and women are estimated for the southern European countries, that is Greece, Italy, Portugal and Spain, with data drawn from 2011 EU-SILC. In Figures 1 and 2, a general picture of the educational and occupational distributions of both men and women is presented. As we can see, gender differentiation in particular occupational classes is evident, while educational distributions show a more similar pattern between sexes.

In Table 3, the distance and similarity measures are presented by country, based on the occupational status. We observe that in general the distance measures indicate low but not negligible rates of gender segregation. Particularly, the higher values of all distance measures are recorded in Portugal, where the

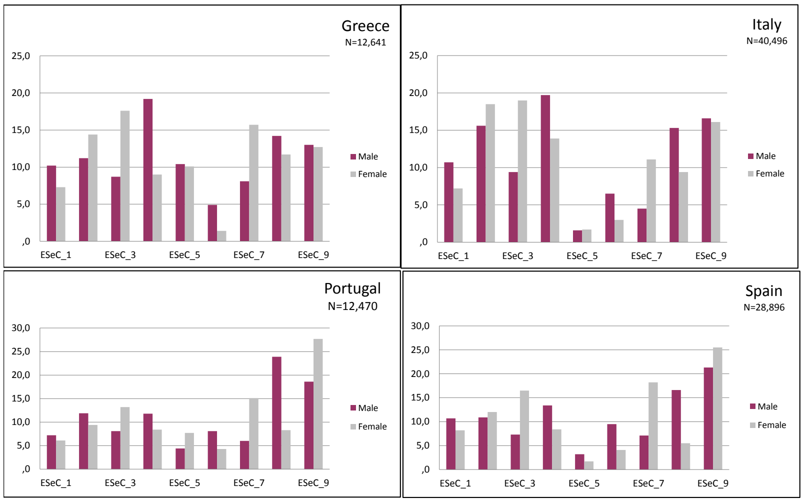


Fig. 1. Occupational distributions between men and women, by country

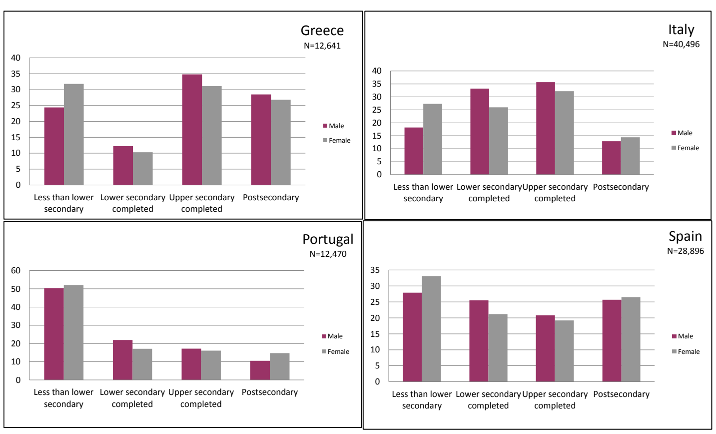


Fig. 2. Educational distributions between men and women, by country

shortest distance, i.e. the Euclidean, equals to 0.219, while the Manhattan distance is 0.530. Italy appears to have the lowest levels of gender differentiation, as can be seen from the very low rates of distance measures.

Country	d_{Euc}	d_{Man}	d_{Sor}	d_{Gow}	d_{Cheb}
Greece	0.167	0.394	0.197	0.044	0.102
Italy	0.154	0.384	0.192	0.043	0.096
Portugal	0.219	0.530	0.265	0.059	0.156
Spain	0.203	0.511	0.255	0.057	0.111

Table 3. Distance and similarity measures based on occupational status of men and women, by country

In the case of Chebychev distance, where only the maximum difference is estimated, the results appear similar among countries, but they are based on different occupational classes. For example, in Greece the maximum difference between men and women is appeared in the 4th occupational class, the small employers and self-employed (non-agriculture), while in Portugal the Chebychev distance shows a drop in the percentage of individuals in the lower technical occupations.

Regarding the case of educational level, in Table 4 the educational distributions between men and women show even lower levels of gender differentiation in all surveyed countries. The highest values are appeared in Italy, while Portugal seems to be the least segregated. It is worth noting that while in other southern countries the maximum difference between sexes appeared in the first educational level, in Portugal the Chebychev distance shows the differentiation in the fourth educational level, where females' participation in post-secondary education is higher than that of males.

Country	d_{Euc}	d_{Man}	d_{Sor}	d_{Gow}	d_{Cheb}
Greece	0.087	0.147	0.073	0.037	0.074
Italy	0.122	0.213	0.106	0.053	0.091
Portugal	0.067	0.118	0.059	0.029	0.048
Spain	0.070	0.119	0.059	0.030	0.052

Table 4. Distance and similarity measures based on educational status of men and women, by country

Table 5 shows the ratio index of gender segregation (R) in the surveyed countries, as well as the $exp(R)$, which measures the overrepresentation of women or men in average occupation. In general, data show medium gender segregation in the southern european labour markets. In Spain, where the highest values of ratio appeared, men and women are overrepresented by the factor of 1.56 in the average occupation, while in Italy, the factor equals to 1.31. The specific ratio R_i, which shows the levels of gender segregation in particular occupational classes, indicates that in all countries men dominate in 1st class (Large employers, higher grade professionals), 4th class (Small employers and self-employed, non-agriculture), 6th class (Lower supervisors and technicians), as well as lower technical occupations, while women are overrepresented in the remaining occupational classes.

Country	**Ratio Index**		**Ration Index for Particular Class Categories**								
	R	$exp(R)$	1	2	3	4	5	6	7	8	9
Greece	0.35	1.42	-0.23	0.36	0.81	-0.65	0.08	-1.14	0.77	-0.09	0.08
Italy	0.27	1.31	-0.37	0.19	0.73	-0.33	0.08	-0.75	0.92	-.047	-0.01
Portugal	0.36	1.44	-0.16	-0.23	0.50	-0.33	0.57	-0.63	0.93	-1.05	0.41
Spain	0.45	1.56	-0.12	0.24	0.96	-0.33	-0.49	-0.70	1.08	-0.96	0.32

Table 5. Ratio Index of Occupational Gender Segregation across South Europe

Regarding the educational level, the equal distributions of men and women in all southern countries observed above are verified, as both R and $exp(R)$ show almost no gender segregation. Women are overrepresented in primary education and in post-secondary education (except for Greece), but the results appeared quite similar for both sexes.

Country	**Ratio index**		**Ratio index scores**			
	R	$exp(R)$	1	2	3	4
Greece	0.03	1.03	0.28	-0.15	-0.09	-0.04
Italy	0.06	1.06	0.36	-0.29	-0.15	0.07
Portugal	0.04	1.05	0.02	-0.26	-0.08	0.32
Spain	0.02	1.02	0.19	-0.17	-0.06	0.05

Table 6. Ratio Index of Educational Gender Segregation across South Europe

5 Conclusions

In this paper, a cross-national comparison of levels of gender segregation was provided between Southern European countries. The main purpose was to show whether and to what extent there is a differentiation in the educational and occupational attainments of men and women, using data drawn from the

2011 European Statistics on Income and Living Conditions. Using the concept of distance and similarity to capture the relationship between individuals' distributions, we observe that regarding the occupational status, distance and similarity measures have low, but not negligible rates, while the index ratios document medium gender segregation in the southern european labour markets, with the highest value corresponding to Spain. Educational distributions between sexes show even lower levels of differentiation, which is also verified by the ratio index.

References

1. G. Betti, A. D'Adostino, L.Neli. Educational mismatch of graduates: A multidimensional and fuzzy indicator. *Social Indicators Research*, 103, 465-480, 2011.
2. M. Charles & D.B. Grusky. Models for Describing the Underlying Structure of Sex Segregation. *American Journal of Society*, Vol. 100, 4:931-971, 1995.
3. F. A. Cowell. Measures of distributional change: an axiomatic approach. *Review of Economic Studies*, 52:135-151, 1985.
4. M. Frechet. Sur quelques points du calcul fonctionnel (These). *Rend. Circ. Mat. Palermo*, 22:1-74, 1906.
5. Gan, G., Ma, C. and Wu, J. (2007), Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, PA.
6. F. Hausdorff. *Grundzuege der Mengenlehre*. Leipzig, 1914. Viet.
7. C. Moser & G. Kalton. *Survey methods in social investigation*. Heineman Educational Books, London, 1971.
8. Valentova M., Krizova I., Katrnak T. Occupational Gender Segregation in the light of the Segregation in Education : A Cross-National Com- parison by Occupational Gender Segregation in the light of the Segregation in Education : A Cross-National Comparison.em IRISS Working Paper Series, 2007.
9. M. Symeonaki & G. Stamatopoulou. Intergenerational mobility as a distance measures between probability distribution functions. *Theoretical and Applied Issues in Statistics and Demography* (C. H. Skiadas (Ed.)), 2015 (to appear).
10. M. Symeonaki, G. Stamatopoulou, C. Michalopoulou. Intergenerational educational mobility in Greece: Transitions and social distances. *Communications in Statistics - Theory and Methods*, 2015 (to appear).

IMITATION MODELLING OF PROCESS OF USING RESOURCE WITH LIMITED PERIOD

Natalia V. Stepanova

Altai Economics and Law Institute, Ave Krasnoarmeisky, 108, 656015 Barnaul, Russia
(E-mail: natalia0410@rambler.ru)

Abstract. A systematic study of inventory models incorporated uncertainly and dynamics began in the early 50s. Nowadays a set of stochastic models are available to solve the inventory control problem under various conditions encountered in practice. The aim of the paper is to stabilize of some type supplies systems performance. The feature of the system under consideration is exogenous (i.e., outside our control) input product flow.
Keywords: modelling, resource with limited period.

A systematic study of inventory models incorporated uncertainly and dynamics began in the early 50s from the works by Arrow, Harris, and Marschak [1] and Dvoretzky, Kiefer, and Wolfowitz [2]. Nowadays a set of stochastic models are available to solve the inventory control problem under various conditions encountered in practice, for examples see Ross [3], Chopra and Meindl [4], and Beyer et al. [5]. The aim of the paper is to stabilize of some type supplies systems performance. The feature of the system under consideration is exogenous (i.e., outside our control) input product flow.

For effective management of industrial or commercial enterprise using a resource with limited period of validity it is necessary to have reliable appraisals of basic characteristics of casual process of employing this resource. These characteristics include mathematical expectations (average values) and dispersions of the following random values: volume of demands for a resource during production cycle X and time of using a lot of resource t. The result of the work of an enterprise can be achievement of statistical (selective) data of two types: 1) real time of using a lot of resource (if there isnt enough resource), 2) real amount of employed resource (if there is some resource left). Natural conditions of the work of an enterprise taken into account, in each of these cases a regular production cycle is over.

The paper examines achievement of statistical estimates for mathematical expectation $M(X)$ and standard deviation $o(X)$ on the basis of combined selection of data of the first type and the second type.

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

These estimates are obtained with the help of two methods: method of moments and method of probability. In order to obtain these estimates asymptotic normality of distributing random values X and t was used.

On the other hand, since the data of the first type and the second type were obtained under the above-mentioned limiting conditions their direct usage for getting statistical estimates of random values X and t is impossible.

In order to get full statistical information about these random values and estimate the accuracy of obtained asymptotic correlations this paper examines imitation modeling of process of using resource with limited period of validity. The use of imitation modeling for additional analysis of the built mathematical model allows us to get volumetric selections which increase the reliability of statistical estimates of the model parameters. Besides, it gives an opportunity to get additional statistical information which is inaccessible under practical obtainment of data. It is achieved with the help of using the notions of fictitious time and fictitious inquiries described below.

Suppose an enterprise buys a lot of resource with volume Q_0, which is to be used during production cycle with duration . The random process $x(t)$, that is, volume of demands within time t is analyzed.

In building mathematical model [6] the following values and suppositions were used:

- demands for a resource come in independently, and value of resource ξ is random value with $M(\xi) = a_1$ and $M(\xi^2) = a_2$;

- the number of demands n $(n \gg 1)$ within time T is random value with $M(n) = m_T$ and $D(n) = \sigma_T^2$; if a flow of demands for a resource is a permanent Poisson flow of intensity λ, then $m_T = \lambda T$, $\sigma_T^2 = \lambda T$.

Under these conditions process $x(t)$ can be approximately described by the diffusive equation

$$dx(t) = m_0 dt + \sigma_0 dw_t; x(0) = 0,$$

where $w(t)$ is standard Wiener process, $m_0 \sim \frac{a_1 m_T}{T} \sim a_1\lambda$, $\sigma_0^2 \sim \frac{m_T(a_2 - a_1^2) + \sigma_T^2 a_1^2}{T} \sim a_2\lambda$ under big values T. Suppose X is a general volume of demands within time T; t is the duration of using the whole lot of resource with volume Q_0.

Within the limits of this model the following results were obtained:

1) Random value X has asymptotically normal distribution: $X \sim N(m_x; \sigma_x^2)$, where $m_x = a_1 m_T = a_1\lambda T$, $\sigma_x^2 = m_T(a_2 - a_1^2) + \sigma_T^2 a_1^2 = a_2\lambda T$.

2) Random value t has asymptotically normal distribution: $t \sim N(m_t; \sigma_t^2)$, where $m_t = Q_0 T/m_x$, $\sigma_t^2 = (\sigma_x^2 Q_0 T^2)/m_x^3$.

3) Estimates of values t and σ_x^2 can be defined by the selection of the type: $x_1, ..., x_i, ..., x_K; t_1, ..., t_j, ..., t_{N-K}$. The method of getting the selection is presented below.

For imitation modeling of this process on the basis of discrete-eventful approach it is necessary to do the following:

1. To generate a flow of demands for a resource.
2. To generate a value of each demand for a resource.

The modeling of a flow of demands for a resource comes to building a sequence of random time moments of coming demands t_j. For Poisson flow of events of intensity λ , the lengths of time intervals between flow events

$\tau_1 = t_1 - t_0, \tau_2 = t_2 - t_1, ...\tau_j = t_j - t_{j-1}, ...$ are independent random values with exponential distribution: $p(\tau) = \lambda e^{-\lambda\tau}(\tau \geq 0)$. The imitation of values of exponential random value τ is obtained with the help of inverse function method: $\tau = -\frac{ln\alpha}{\lambda}$, where α is a random value which has uniform distribution in $(0;1]$. Thus, $t_j = t_{j-1} - \frac{ln\alpha}{\lambda, t_0=0}$. For each moment t_j a random value of demand value is generated, that is the realization ξ_j of random value ξ with mathematical expectation $M(\xi) = a_1$, and dispersion $D(\xi) = a_2 - a_1^2$.

The condition of terminating modeling process is condition $x \geq Q_0$ or $t \geq T$.

The realization of process $x(t)$ is shown graphically in Figure 1.

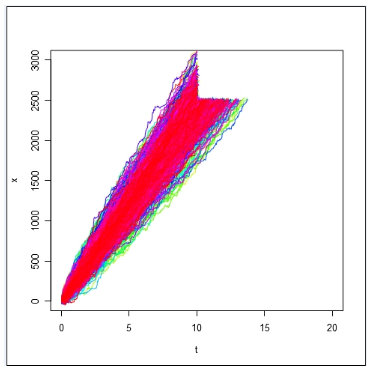


Fig. 1. The realization of process $x(t)$, obtained as a result of imitation modeling, $N = 1000, a_1 = 8, a_2 = 75, \lambda = 30, Q_0 = 2400, T = 10$

Suppose the result of imitation modeling is N realizations of random process $x_1(t), x_2(t), ...x_N(t)$. There may be two cases for each realization:

1) Resource in number $X < Q_0$, was used at the end of the production cycle, so the resource remained unused.

2) The use of the resource finished at moment $t < T$, so there was not enough resource.

To get selective data according to the model we consider fictitious time in the first case and fictitious demands for a resource in the second case. It means the following:

1) Fictitious time is a period of time after the moment of terminating a production cycle up to the moment of using the whole lot of resource.

2) Fictitious demands for a resource are demands which came in after the moment of termination of resource up to the end of production cycle.

Suppose for some realization $x_i(t)$ of process $x(t)$ condition 1 is observed. In this case let us designate by $x_i = x_i(T)$ the amount of resource used during the production cycle and by t_i the period of time during which the whole lot of resource (taking into account fictitious time) will be completely used. If for some realization $x_j(t)$ condition 2 is observed, then t_j is termination moment of using a resource and $x_j = x_j(T)$ is the general volume of demands during time T (taking into account fictitious demands).

We get selective data which will be used the following way:

1) selection $x_1, x_2, x_3, ..., x_N$ is for estimating numerical characteristics and checking the hypothesis about a normal type of distribution for random value X, general volume of demands for a resource during time T.

2) selection $t_1, t_2, t_3, ..., t_N$ is for estimating numerical characteristics and checking the hypothesis about a normal type of distribution for random value t time length of using the whole lot of resource.

3) selection $x_1, ..., x_i, ..., x_K; t_1, ..., t_j, ..., t_{N-K}$ (after the change of data numeration), in which values x_i are chosen in the first case and values t_j in the second case. This selection will be used for estimating input characteristics of the model by the method of maximum probability.

Suppose $\bar{m_x}, \bar{\sigma}_x^2$ are estimates of mathematical expectation and dispersion of random value X; and $\bar{m_t}, \bar{\sigma}_t^2$ are estimates of random value t. To estimate mathematical expectation of random values X and t we use the selective means: $\bar{m_x} = \frac{1}{N}\sum\limits_{i=1}^{N} x_i$, $\bar{m_t} = \frac{1}{N}\sum\limits_{i=1}^{N} t_i$; dispersions X and t are estimated with the help of unchanged selective dispersions: $\bar{\sigma}_x^2 = \frac{1}{N-1}\sum\limits_{i=1}^{N}(x_i - m_x)^2$, $\bar{\sigma}_t^2 = \frac{1}{N-1}\sum\limits_{i=1}^{N}(t_i - m_t)^2$.

The verification of statistical hypotheses about a type of distribution has been conducted with the help of Pearsons criterion of agreement χ^2. The main hypothesis $H_0 : t \in N(\bar{m_x}, \bar{\sigma}_x^2)$ is checked against the alternative hypothesis $H_1 : t \notin N(\bar{m_x}, \bar{\sigma}_x^2)$ with significance level α. The calculation of distribution parameters, observed value criterion and probability is made with the help of the program StatSoft Statistica v6.0 by selective values $x_1, x_2, x_3, ..., x_N$. In the same way the check of the hypothesis $H_0 : t \in N(\bar{m_x}, \bar{\sigma}_x^2)$ is realized against the alternative hypothesis $H_1 : t \notin N(\bar{m_x}, \bar{\sigma}_x^2)$ by alternative data $t_1, t_2, t_3, ..., t_N$.

The comparison of estimates of mathematical expectation and standard deviation of random value X, which were obtained by two ways: 1) $(m_x; \sigma_x^2)$ by direct selection $x_1, x_2, x_3, ..., x_N$; 2) $(\hat{m}_x; \hat{\sigma}_x^2)$ with the help of method of maximum probability by selection $x_1, ..., x_i, ..., x_K; t_1, ..., t_j, ..., t_{N-K}$ is given in Table 1.

N	1000	1000
T	12	9
Q_0	2400	900
$\bar{m}_x$	2401,19	909,96
$\hat{m}_x$	2400,78	867,96
$\lvert\hat{m}_x - \bar{m}_x\rvert$	0,41	42,84
$\lvert\hat{m}_x - \bar{m}_x\rvert/\bar{m}_x$	0,015	4,71
$\bar{\sigma}_x$	57,75	148,32
$\hat{\sigma}_x$	59,43	131,80
$\lvert\hat{\sigma}_x - \bar{m}_x\rvert$	1,68	16,52
$\lvert\hat{\sigma}_x - \bar{m}_x\rvert/\bar{m}_x$	2,91	11,13

Table 1.

The comparison of estimates of mathematical expectation and dispersion of random value t, which were calculated by two ways: 1)$(\bar{m}_t; \bar{\sigma}_t^2)$ by direct selection $t_1, t_2, t_3, ..., t_N$; 2) $(\hat{m}_t; \hat{\sigma}_t^2)$ with the use of theoretical correlations $\hat{m}_t = Q_0 T/\bar{m}_t$, built as a result of three experiments of imitation modeling of realizing random process $x(t)$, is given in Table 2.

	1	2	3	4
N	1000	1000	400	1000
T	12	10	9	9
Q_0	2400	2500	1000	900
$\bar{m}_x$	2401,19	2405,05	1086,48	909,96
$\bar{\sigma}_x$	57,74	241,17	60,41	148,32
$\bar{m}_t$	12,0	10,46	8,33	9,044
$\hat{m}_t$	11,99	10,39	8,28	8,90
$\|\hat{m}_t - \bar{m}_t\|$	0,01	0,07	0,05	0,144
$\|\hat{m}_t - \bar{m}_t\|/\bar{m}_t$	0,1	0,7	0,6	1,6
$\bar{\sigma}_t$	0,084	1,025	0,243	2,271
$\hat{\sigma}_t$	0,083	1,045	0,230	2,128
$\|\hat{\sigma}_t^2 - \bar{\sigma}_t^2\|$	0,001	0,02	0,013	0,143
$\|\hat{\sigma}_t^2 - \bar{\sigma}_t^2\|/\bar{\sigma}_t^2$	1,3	2,0	5,4	6,3

Table 2.

The results of numerical experiments considered in this paper are the following:

1. They statistically prove the obtained in [6] theoretical computations regarding normality of distributing a process of demands for a resource;

2. They show a good coincidence of estimates obtained by the observed in the process of using a resource combined selection $x_1, ..., x_i, ..., x_K; t_1, ..., t_j, ..., t_{N-K}$ with the help of method of maximum probability [6] with standard selective estimates of average and dispersion, which were obtained by selections with addition of fictitious observations which are not accessible in practice.

References

1. K.J. Arrow, Th.E. Harris, J. Marschak. Optimal Inventory Policy. *Econometrica*, 3, **19**, 250–272, 1951.
2. A. Dvoretzky, J. Kiefer, J. Wolfowitz. On the optimal character of the (S, s) policy in inventory theory. *Econometrica*, 4, **21**, 586–596, 1953.
3. Sh.M. Ross *Applied probability models with optimization applications*, New York: Dover Publications, 1992.
4. S. Chopra, P. Meindl. *Supply chain management*, London: Prentice Hall, 2001.
5. D. Beyer, F. Cheng, S.P. Sethi and M. Taksar. *Markovian demand inventory models.*, New York: Springer, 2010.
6. A.V. Kitaeva, N.V. Stepanova. Control of Optional Resources. *Tomsk Polytechnic University Bulletin*, 5, **322**, 23–28, 2013.

Identifying Quality of Life patterns in the third age: a cluster analysis approach with evidence from SHARE

Maria Symeonaki

Department of Social Policy, School of Political Sciences, Panteion University of Social and Political Sciences, Athens, Greece
(E-mail: msymeon@panteion.gr, msymeonaki@me.com)

Abstract. The observed increase of life expectancy in Europe has naturally led to a literature growth on aspects that can influence the quality of life of the elderly. The measurement of quality of life has therefore become a significant issue, emphasising on the features that are important for establishing and guaranteeing high levels of quality of life. The present paper contributes to the rising number of studies on measuring quality of life by considering a cluster analysis approach on the SHARE databases[1] to further analyse and specify patterns of quality of life in European countries. The SHARE-project is a multidisciplinary, longitudinal, and cross-national panel database, developed to understand the relations between health, labour force participation, and institutional context of old people support in Europe. The cluster analysis outcomes are combined with the results of the analysis of the scale used to measure quality of life, i.e. the CASP-12 scale and different patterns of quality of life are identified using the cluster centres.
Keywords: Cluster Analysis, Quality of Life, CASP-12 scale.

1 Introduction

The measurement of quality of life has become in recent years an issue of utmost importance, emphasizing on the aspects that are significant for establishing and ensuring a high quality of life. The observed increase of life expectancy in Europe has reasonably led to a literature growth on aspects that

[1] This paper uses data from SHARE wave 4 release 1, as of November 30th 2012.
The SHARE data collection has been primarily funded by the European Commission through the 5th Framework Programme (project QLK6-CT-2001-00360 in the thematic programme Quality of Life), through the 6th Framework Programme (projects SHARE-I3, RII-CT-2006-062193, COMPARE, CIT5- CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th Framework Programme (SHAREPREP, No 211909, SHARE-LEAP, No 227822 and SHARE M4, No 261982). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, R21 AG025169, Y1-AG-4553-01, IAG BSR06-11 and OGHA 04-064) and the German Ministry of Education and Research as well as from various national sources is gratefully acknowledged (see www.shareproject.org for a full list of funding institutions).
This study was supported by a grand from THALIS-Panteion University-Investigating Crucial Interdisciplinary Linkages in Ageing Societies.

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

can influence the quality of life of the elderly (see for example Hyde et al.[6], Knesebeck et al.[7], Wahrendorf[11], Young and Schuller[13], Motel-Klingebiel et al.[8], Motel-Klingebiel et al.[9], Crimmins et al.[4], among others). Measuring quality of life can be subjective, but is nonetheless a very useful complement to more objective data when comparing well-being across countries.

Aiming to contribute to the above discussion, in the present paper we investigate the measurement of quality of life for European countries by providing patterns of quality of life. The measurement is based on an indicator specifically targeted at older populations, the CASP-12 scale, which measures the degree to which the old person has his/her need covered and which is actually a short form of the CASP-19 scale. CASP-19 is a subjective measure of well-being derived from an explicit theory of human need, spanning four life domains: control, autonomy, self-realisation and pleasure and is a self-reported summative index consisting of 12 Likert scale items (Wiggins et al.[12], p.61).

The main aim of this paper is twofold:

- to capture the differences between European countries, taking also into account a North-South and East-West gradient, and
- to provide patterns of quality of life among European countries using cluster analysis.

Our analysis is based on the latest available data drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE-project), i.e. data drawn from wave 4 (2010). Unfortunately, data from wave 5, whose fieldwork was completed in November 2013, is not available yet.

The main benefits from using the SHARE databases is that they provide comparable information for participating European countries, concerning a very wide variety of aspects, such as quality of life, mental health, labour force participation, working conditions, quality of work, early retirement, etc. Moreover, the data is available to the scientific community free of charge after registration.

The present paper is organized in the following way. Section 2 presents the data used, whereas Section 3 deals with measuring quality of life for wave 4 in SHARE, for all participating countries. Section 4 provides the clustering of respondents and presents the outcomes of the analysis. A map of low, medium and high quality of life across countries for wave 4 is produced and the socio-demographic characteristics of respondents belonging to all clusters are provided. The conclusions of the analysis are discussed in Section 5.

2 Data

In order to identify patterns of quality of life, we use data drawn from the Survey of Health, Ageing and Retirement in Europe (SHARE-project). The SHARE project (Börsch-Supan and Jürges[3]) is a multidisciplinary, longitudinal and cross-national panel database, developed to understand the relations between health, labour force participation and institutional context of old people support in Europe. Funded mainly via the European Commission, as well as the US National Institute on Ageing and national sources, it was firstly designed in January of 2002 and it is conducted every two years. The first wave of the survey took place in 2004-2005, in 11 European countries ranging from Nordic to Mediterranean countries, while the fifth wave took place in 2012-2013. The main purpose of the survey is to provide a full picture of all aspects of ageing process and its impact in the different cultures of Europe. Moreover, using the knowledge of its predecessors, the US Health and Retirement Study (HRS) and the English Longitudinal Survey on Ageing (ELSA), the survey aims to collect comparable data useful for the policies planned and applied in the European Union. In order to succeed its purposes, the survey was divided into 21 modules. Except for the coverscreen (CV) and the demographic (DN) modules, it covers a large variety of subjects, such as physical and mental health, behavioral risks, employment and pensions, social support etc. The target population of the survey is all the non-institutionalized population aged more than 50 years old, as well as their spouse, regardless of their age. Note that wave 3 in SHARE corresponds to a different survey, SHARELIFE, which is a retrospective survey on people's life history. The total sample size in each participating country for waves 1, 2 and 4 are presented in Table 1.

Table 1. Number of respondents in SHARE countries, Waves 1, 2, 4

Country	**Wave 1**	**Wave 2**	**Wave 4**
Austria	1,893	1,341	5,286
Belgium	3,827	3,169	5,300
Czechia		2,830	6,118
Denmark	1,707	2,616	2,276
Estonia			6,828
France	3,193	2,968	5,857
Germany	3,008	2,568	1,572
Greece	2,898	3,243	
Hungary			3,076
Ireland		1,134	
Israel	2,598		
Italy	2,559	2,983	3,583
Netherlands	2,979	2,661	2,762
Poland		2,467	1,724
Portugal			2,080
Slovenia			2,756
Spain	2,396	2,228	3,570
Sweden	3,053	2,745	1,951
Switzerland	1,004	1,462	3,750
All countries	31,115	34,415	58,489

Source: http://www.share-project.org/data-access-documentation/sample.html

3 Measuring quality of life (CASP-12)

In Hyde et al.[6] a measure of subjective quality of life targeted at older populations, the CASP scale, was proposed. Four constructs, referring to conceptual domains of individual needs that are important in the early old age, i.e. control (C), autonomy (A), self-realisation (S) and pleasure (P) form the scale, originally comprised of 19 items. A revised psychologically validated form of 12 items is often used as it shows stronger measurement properties in its reduced form (Wiggins et al.[12], p. 75). SHARE includes this revised form, the CASP-12 questionnaire and the respective questions are presented in Table 2. For each of the constructs, three questions are asked, and each one is rated by an ascending 1 to 4 scale. Therefore, the total score of the indicator records values that range from 12 to 48. The relevant variables can be found in the Activities (AC) module.

Table 2. Quality of Life: CASP-12

Here is a list of statements that people have used to describe their lives or how they feel. We would like to know how often, if at all, you think they apply to you.	SHARE-variable (1=often, 4=never)
Control	
My age prevents me from doing the things I would like to.	ac014
I feel that what happens to me is out of control.	ac015
I feel left out of things.	ac016
Autonomy	
I can do the things that I want to do.	ac017*
Family responsibilities prevent me from doing what I want to do.	ac018
Shortage of money stops me from doing the things I want to do.	ac019
Pleasure	
I look forward to each day.	ac020*
I feel that my life has a meaning.	ac021*
On balance, I look back on my life with a sense of happiness.	ac022*
Self-realisation	
I feel full of energy these days.	ac023*
I feel that life is full of opportunities.	ac024*
I feel that the future looks good for me.	ac025*

** Items reverse coded for scoring, AC module*

To examine whether there are specific patterns of quality of life we first estimate the sum scores among all countries. The values of all positively worded items are reversed in order to achieve correspondence between the ordering of the response categories. Table 3 presents the CASP-12 scale's Cronbach's alpha for all countries participating in SHARE and the mean CASP-12 scores by country and gender for wave 4.

It can easily be seen that quality of life scores are comparatively low for Italy (rank 16), Chechia (rank 15) and Poland (rank 14) and comparatively high for Switzerland (rank 1), Denmark (rank 2) and the Netherlands (rank 3).

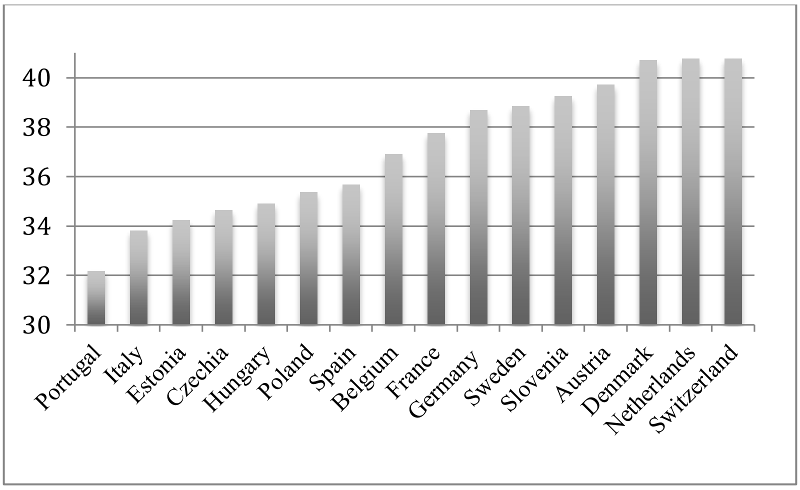
Differences between countries are highly significant at p<0.001, (Independent Samples, Kruskal-Waliis Test).

In Figure 1 the mean scores of quality of life by country are presented for wave 4 and Figure 2 provides the mean CASP-12 scores by country and gender.

Table 3. Cronbach's alpha and mean CASP-12 scores by country and gender, wave 4

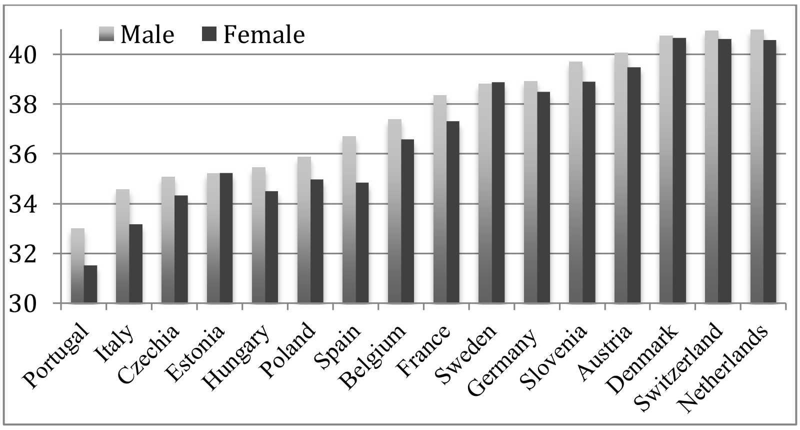
Country	**Cronbach's alpha**	**Mean CASP-12, Total**	**Mean CASP-12, Males**	**Mean CASP-12, Females**
Austria	0.822	39.71	40.05	39.47
Germany	0.769	38.68	38.90	38.48
Sweden	0.771	38.84	38.81	38.87
Netherlands	0.736	40.75	40.98	40.57
Spain	0.810	35.67	36.69	34.83
Italy	0.769	33.80	34.57	33.17
France	0.792	37.74	38.34	37.30
Denmark	0.767	40.70	40.75	40.66
Switzerland	0.767	40.76	40.95	40.61
Belgium	0.763	36.90	37.37	36.57
Czechia	0.787	34.63	35.06	34.32
Poland	0.841	35.36	35.87	34.96
Hungary	0.799	34.90	35.44	34.50
Portugal	0.686	32.16	33.00	31.51
Slovenia	0.780	39.25	39.70	38.90
Estonia	0.807	34.22	35.20	35.23

Source: SHARE-project: sharew4_rel2-6-0_ac.sav



Source: SHARE-project: sharew4_rel2-6-0_ac.sav

Fig. 1. Mean quality of life scores by country, wave 4



Source: SHARE-project: sharew4_rel2-6-0_ac.sav

Fig. 2. Mean quality of life scores by country and gender, wave 4

It is apparent that significant differences exist between countries in respect to the measurement of quality of life. Moreover, in all countries except for Estonia, Sweden and Austria women seem to score lower than men on the CASP-12 scale. The hypothesis testing (Mann-Witney U test) showed that distributions are the same for Sweden, the Netherlands, Denmark and Estonia and the medians are the same for Sweden, the Netherlands, Denmark, Switzerland, Poland, Estonia and Germany.

3 Clustering the respondents

In order to examine the geographic variation of quality of life among European countries we first divide quality of life scores into 3 classes of low, medium and high quality of life and create the map presented in Figure 3, using a Geographical Information System (GIS) software. Apparently, the results confirm a North-South gradient except for Spain that scores higher than France for example.

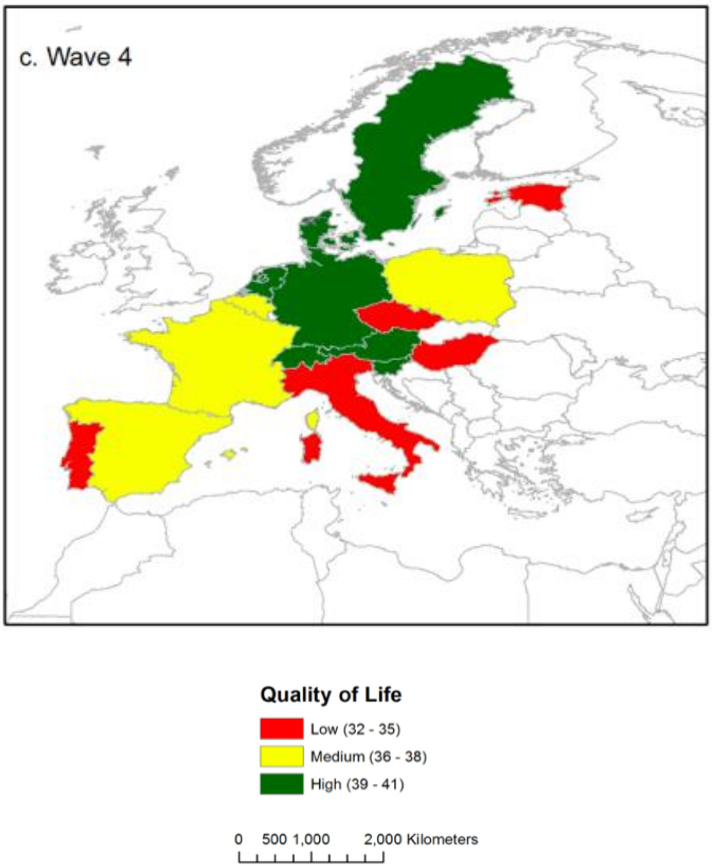


Fig. 3. Map of quality of life, wave 4

Now, the next step is to cluster the respondents of the above mentioned categories of low quality of life (QoL), medium QoL and high QoL, using K-means clustering. In general, clustering involves the task of dividing data points into clusters so that the items (in our case respondents) in the same cluster are as similar as possible and items in different classes are as dissimilar as possible.

Clustering techniques are used in order to convert a large number of items into a small number of representative clusters, using the cluster centres as prototypes (Everitt[5], Aldenderfer and Blashfield[1]). K-means is an unsupervised learning algorithm that classifies a given data set through a specific number of clusters fixed a priori. Here, we use a fixed number of clusters equal to four that corresponds to the response categories of each item of the CASP-12 scale. The mean CASP-12 scores per QoL categories are presented in Table 4.

Table 4. Mean CASP-12 scores per QoL categories

Clusters per spatial category	**mean QoL**
Low Quality of Life (Italy, Portugal, Estonia, Czechia, Hungary)	
Cluster 1	22.39
Cluster 2	29.58
Cluster 3	35.92
Cluster 4	42.69
Medium Quality of Life (Poland, Spain, Belgium, France, France)	
Cluster 1	23.30
Cluster 2	30.51
Cluster 3	37.15
Cluster 4	43.60
High Quality of Life (Germany, Sweeden, Slovenia, Austria, the Netherlands, Switzerland, Denmark)	25.26
Cluster 1	32.74
Cluster 2	38.81
Cluster 3	44.50
Cluster 4	

Obviously, the first cluster corresponds to respondents with lower scores and Cluster 4 to respondents with higher ones.
The demographic and social characteristics of the members of each cluster are presented in Tables 5, 6 and 7. In order to measure the health state of the respondents the ph002 variable from the Physical Health (PH) module was used.

In all QoL categories a more dissatisfied attitude represented by the first cluster would appear among older, female respondents with lower education and poorer health. On the other hand highly satisfied attitudes (Cluster 4) would appear among those aged between 50 and 64, who are of a higher education and have better health in all QoL categories.

Table 5. The characteristics of respondents belonging to low QoL countries by cluster

Low Quality of Life Countries				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Age Classes				
50-64	39.4	50.5	59.3	65.0
65-74	31.2	31.2	28.9	26.1
75+	29.4	18.3	11.8	8.9
Education				
Low	35.2	26.3	16.0	8.3
Medium	59.2	66.0	71.3	72.7
High	5.6	7.7	12.7	19.0
General Health				
Poor	55.6	25.2	10.6	3.6
Fain	7.8	24.3	36.1	40.6
Good	36.0	49.0	50.5	48.0
Excellent	0.5	1.5	2.9	7.7
Gender				
Males	34.9	39.7	44.2	43.7
Females	65.1	60.3	55.8	56.3

Table 6. The characteristics of respondents belonging to medium QoL countries by cluster

Medium Quality of Life Countries				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Age Classes				
50-64	38.7	50.3	57.1	65.2
65-74	24.6	25.6	25.9	23.4
75+	36.7	24.1	17.0	11.4
Education				
Low	35.8	24.5	16.2	9.9
Medium	58.0	65.1	69.5	69.7
High	6.2	10.4	14.3	20.4
General Health				
Poor	49.8	21.2	6.9	1.6
Fain	12.8	31.7	44.1	46.9
Good	37.2	45.5	44.1	39.8
Excellent	0.2	1.5	4.8	11.6
Gender				
Males	31.0	40.9	44.7	48.3
Females	69.0	59.1	55.3	51.7

Table 7. The characteristics of respondents belonging to high QoL countries by cluster

Low Quality of Life Countries				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Age Classes				
50-64	38.6	46.7	54.5	60.4
65-74	29.5	29.6	29.9	28.8
75+	31.9	23.7	15.6	10.8
Education				
Low	36.2	34.6	32.7	30.3
Medium	57.2	57.9	56.5	58.1
High	6.5	7.6	10.7	11.6
General Health				
Poor	41.9	15.8	5.2	1.5
Fain	15.9	32.2	39.7	35.7
Good	41.0	48.8	47.7	44.6
Excellent	1.2	3.3	7.4	18.2
Gender				
Males	38.9	41.3	44.6	45.1
Females	61.1	58.7	55.4	54.9

4 Conclusions and Discussion

In the present paper we have used clustering techniques (K-means) to identify patterns of quality of life. From the analysis, it follows that in general older women, less educated and of poorer health are more dissatisfied with their lives. Moreover, there is evidence that a North-South gradient exists, since differences between countries are highly significant and northern European countries (Austria, Denmark, the Netherlands, Sweden, etc) exhibit by far the highest scores. However, age, gender, education and health condition seem to play an important role in all gradients of quality of life.

References

1. M. Aldenderfer and R. Blashfield. Cluster Analysis (Quantitative Applications in the Social Sciences), Sage, London, 1984.
2. A. Börsch-Supan, A. Brugiavini, H. Jürges, A. Kapteyn, J. Mackenbach, J. Siegrist, and G. Weber. First results from the Survey of Health, Ageing and Retirement in Europe (2004-2007). Starting the longitudinal dimension. Mannheim: Mannheim Research Institute for the Economics of Aging (MEA), 2008.
3. A. Börsch-Supan and H. Jürges, (eds.). The Survey of Health, Ageing and Retirement in Europe - Methodology. Mannheim, Mannheim Research Institute for the Economics of Aging (MEA), 2005.

4. E. Crimmins, K. Jung, and S.-A. Aida. Gender differences in health: results from SHARE, ELSA and HRS. European Journal of Public Health 1(1), 81-91, 2010.
5. B. Everitt, Cluster Analysis, Arnold, New York, 1993.
6. M. Hyde, R. Wiggins, P. Higgs and D. Blane. A measure of quality of life in early age: the theory, development and properties of a needs satisfaction model (CASP-19). Aging and Mental Health, 7, 186-194, 2003.
7. O. von dem Knesebeck, M. Wahrendorf, M. Hyde, and J. Siegrist. Socio-economic position and quality of life among older people in 10 European countries: Results of the SHARE study. Ageing and Society 27, 269-284, 2007.
8. A. Motel-Klingebiel, H.-J. von Kondratowitz and C. Tesch-Römer. Social Inequality in the Later Life: Cross-National Comparison of Quality of Life. European Journal of Ageing, 1, 6-14, 2004.
9. A. Motel-Klingebiel, G. Laura Romeu, B. Jorg. Welfare states and quality of later life: distributions and predictions in a comparative perspective. European Journal of Ageing, 6, 67-78, 2009.
10. Survey of Health, Ageing and Retirement in Europe (SHARE), Manheim. Available at http://www.share-project.org/ (29 March 2015, data last accessed).
11. M. Wahrendorf, O. von dem Knesebeck and J. Siegrist. Social productivity and well-being of older people: Baseline results from the SHARE study. European Journal of Ageing, 3, 67-73, 2006.
12. R. Wiggins, G. Netuveli, M. Hyde, P. Higgs, D. Blane. The evaluation of a self-enumerated scale of quality of life (CASP-19) in the context of research on ageing: a combination of exploratory and confirmatory approaches. Social Indicators Research, 89, 61-77, 2008.
13. M. Young and T. Schuller. Life after work: The arrival of the ageless society. London: Harper Collins, 1991.

Strategic maintenance planning by fuzzy AHP and Markov Decision Processes

Sahar Tahvili[1], Jonas Österberg[1], and Sergei Silvestrov[1]

Division of Applied Mathematics, School of Education, Culture and Communication, Mälardalen University, Box 883, SE-721 23 Västerås, Sweden
(e-mail: jonas.osterberg@mdh.se, sti10002@student.mdh.se, sergei.silvestrov@mdh.se)

Abstract. The work of engineering and business professionals includes making a series of decisions and optimizations. Real world decision making problems faced by decision makers (DM) involve multiple, usually conflicting, criteria. These multi-criteria decision making problems (MCDM) are usually complicated and large in scale. In strategic Maintenance planning, choices are made on where to focus time and effort, where to spend money. We consider a framework for strategic maintenance planning in a modern maintenance driven organization. Our focus is on a multi-stage framework in which the planning is divided into two stages, identifying an optimal set of possible actions and finding the optimal decision policy for these actions for each point in time as a function of the stochastically evolving system state. To this respect we consider the MCDM method of AHP (Analytical hierarchical programming) in a fuzzy environment, and Markov decision processes (MDP).
Keywords: AHP, Fuzzy Logic, Markov Decision Process, Maintenance Planing, Optimization, MCDM.

1 Introduction

One of the most important factors in the operations of many companies today is to maximize profit; maintenance management of equipment is an important tool to that effect. Maintenance activities is at the largest level divided into two major areas, preventive maintenance activities (PM) and corrective maintenance activities (CM). Preventive maintenance is performed when the system is operational and to avoid future system failure while corrective maintenance is, per definition, performed as a response to a system failure. Creating maintenance plans with a maintenance strategy consists of data mining, mathematical modeling, optimization and making a series of decisions. Some of the decision making and optimization problems that are typically faced by the decision makers (DM) involve multiple, usually conflicting criteria. Multi criteria decision making problems (MCDM) are complicated and generally of large scale.

Real world optimization problems usually include more than one objective function to be optimized simultaneously. The principal objective of this research is to demonstrate two different approaches for decision making. We

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

introduce TOPSIS and AHP as two well-known techniques for solving MCDM problems and generalize this to fuzzy multi criteria decision-making (FMCDM).

Multi objective optimization problems can be stated in mathematical terms as:

$$\begin{aligned}
&\min f(x) = (f_1(x), ..., f_m(x)) \\
&s.t. \\
&x \in X = \{x \in R^n \geq 0, i = 1, 2, ..., n\}
\end{aligned}$$

where x is an n-dimensional vector of variables which indicates our alternatives and the objective functions $(f_1(x), ..., f_m(x))$ is an m-dimensional vector of the criteria.$g_i(x)$ indicates the functional constraints inflicted by the multi criteria problem [6]. In other words, $m \geq 2$ is the number of objectives and the set X is the feasible set of decision vectors.

2 Previously proposed solution

In our previous work [8], we have introduced a model-based approach for solving a MCDM problem. As part of this approach we have also defined the problem in a Fuzzy environment.

As we mentioned earlier, more often the criteria contains imprecision, inherent vagueness (conflicting criteria) and the data are imprecise and fuzzy. Since we are not able to define MCDM problems by crisp value (ordinary real numbers), we use fuzzy numbers as a suitable alternative which provides a useful way to solve MCDM problems. In other words, by using fuzzy logic we are able to correlate the problem better with the reality.

In our previous work [8], we assumed a decision making situation where every criteria effects every single alternative. Idle Roller (a component in the cooling system of heavy trucks) was chosen as the case study. We identified 15 various criteria (such as temperature, humidity, mileage, etc.) that have direct effect on the health of the Idle Roller.

Imperfect maintenance, corrective maintenance and preventive maintenance were the different maintenance activities proposed as alternatives. Figure 1 represents the previous problem statement.

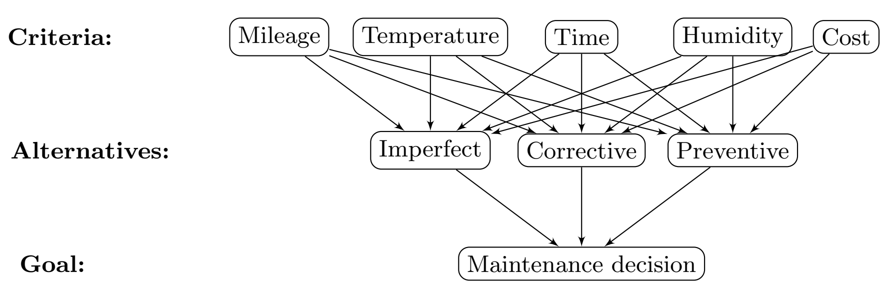


Fig. 1. Problem statement in the previous work solved by Fuzzy TOPSIS

2.1 Fuzzification

Fuzzification consists of the process of transforming the linguistic variables to fuzzy sets [1]. We need to use membership functions to associate a grade to each linguistic term and have an accurate data analysis over the health of system. In our previous work [8] we used RapidMiner as a suitable tool for data mining to measure the causal effect that every criteria have on the system efficiency. We used Fuzzy TOPSIS method, which is a well-known decision making technique, based on two ideal solutions:

$$\tilde{A}^{+} = \{(C_1, \text{Max}\{\mu_{i,1}\}, \text{Min}\{\nu_{i,1}\}), \cdots, (C_m, \text{Max}\{\mu_{i,m}\}, \min\{\nu_{i,m}\})\}.$$
$$\tilde{A}^{-} = \{(C_1, \text{Min}\{\mu_{i,1}\}, \text{Max}\{\nu_{i,1}\}), \cdots, (C_m, \text{Min}\{\mu_{i,m}\}, \text{Max}\{\nu_{i,m}\})\}.$$

where $\tilde{A}^{+}$ and $\tilde{A}^{-}$ indicates the fuzzy positive and fuzzy negative ideal solution respectively.

Definition 1. [11] A fuzzy set is a pair (A, m) where A is a set and $m : A \to [0, 1]$. For each $x \in A, m(x)$ is called the grade of membership of x in (A, m). For a finite set $A = \{x_1, \dots, x_n\}$, the fuzzy set (A, m) is often denoted by $\{m(x_1)/x_1, \dots, m(x_n)/x_n\}$. Let $x \in A$. Then x is called fully included in the fuzzy set (A, m) if $m(x) = 1$ and is called not included if $m(x) = 0$. The set $\{x \in A | m(x) > 0\}$ is called the support of (A, m) and x is a fuzzy member if $0 < m(x) < 1$.

Definition 2. [14] For any set X a membership function on X is any function from X to the real unit interval $[0, 1]$, the membership function which represents a fuzzy set A is denoted by μ_A. For an element x of X, the value $\mu_A(x)$ is called the membership degree of x in the fuzzy set A .

Definition 3. [2] An Intuitionistic Fuzzy Set (IFS) A on a universe U is defined as an object of the following form:
$A = \{(u, \mu_A(u), \nu_A(u)) \mid u \in U\}$, where the functions $u_A : U \to [0, 1]$ and $v_A : U \to [0, 1]$ define the degree of membership and the degree of non-membership of the element $u \in U$ in A, respectively, and for every $u \in U$ we have $0 \leq \mu_A(u) + \nu_A(u) \leq 1$,

According to [13] any fuzzy set can be written as:

$$\{(u, \mu_A(u), 1 - \mu_A(u)) \mid u \in U\}. \tag{1}$$

IFS distribute fuzzy sets for every membership function μ and non-membership functions ν where $\nu = 1 - \mu$.

Definition 4. [12] Assume U is a finite universe and R is a Residual implication. I_{IFS} is an inclusion degree function if:

- $\forall a, b \in [0, 1]$ and $a \leq b \Rightarrow R(a, b) = 1$

- $R(a, b)$ is non-decreasing with respect to b and non-increasing with respect to a.

By using Definition 4 we are able to write:

$$I_{IFS}(A,B) = \frac{1}{\mid U \mid} \sum_{u \in U} [\lambda R(\mu_A(u), \mu_B(u)) + (1-\lambda) R(\nu_B(u), \nu_A(u))], \tag{2}$$

where $\lambda \in [0, 1]$ and $\mid U \mid$ is the cardinality of U which can be calculated by [4]

$$\mid U \mid = \sum_{u \in U} \frac{1 + \mu_A(u) - \nu_A(u)}{2}. \tag{3}$$

There are different methods to calculate an R-implication which was introduced by several mathematicians. we used Lukasiewics implication:

$$R_L(a, b) = \text{Min}\{1 - a + b, 1\}. \tag{4}$$

To find the best alternative we need to measure the geometric distances between every alternative and the positive and negative ideal solutions respectively.

The geometric distances can be calculated with:

$$\tilde{D}^+(M_i) = \text{Max}(I(\tilde{A}^+, M_i)), \tag{5}$$

$$\tilde{D}^-(M_i) = \text{Min}(I(M_i, \tilde{A}^-)) \tag{6}$$

where $\tilde{D}^+(M_i)$ is the fuzzy geometric distance between criteria M_i and $\tilde{A}^+$ and $\tilde{D}^-(M_i)$ is the fuzzy geometric distance between criteria M_i and $\tilde{A}^-$ and I indicates to the inclusion degree.

An optimal alternative which has the shortest distance from the positive ideal solution and the longest distance from the negative ideal solution simultaneously, which is called Ranking index and can be calculated by:

$$p_i = \frac{\tilde{D}^+(M_i)}{\tilde{D}^-(M_i) + \tilde{D}^+(M_i)}, \tag{7}$$

Let $i_0 = \text{argmax}\, p_i$ then the alternative M_{i_0} is best alternative [13].

We summarize here the result of our previous work as:

- The goal: Decision support system for a maintenance planing,
- The technique: Fuzzy TOPSIS,
- The tools: Matlab, RapidMiner,
- Alternatives: Imperfect maintenance, preventive maintenance, corrective maintenance,
- Criteria: Temperature, Humidity, Mileage,Time, Cost, etc.
- The case: Idler Roller.

By using Equations (1), (2), (3), (4), (5) and (7) we find the Preventive Maintenance as the best maintenance activity for Idle Roller on the cooling system.

3 The Analytic Hierarchy Process (AHP)

In the present paper, we introduce and use AHP. For using AHP technique, which is based on a decision matrix [10], we change our assumption and redefine the initial problem in the new situation with different conditions.

Let us assume a new decision making situation where the decision maker wants to choose the best alternative among: corrective maintenance, preventive maintenance, imperfect maintenance and inspection. We use Belt Tensioner in the cooling system as our case study.

Identifying the criteria is the first stage of decision making process. In our previous work [8] we have found 15 criteria which have direct effect on the health of the component.

Since in real decision making situations some criteria have sub-criteria, (for example the cost as a criteria can be divided into three groups: fine cost, maintenance cost and breakdown cost) we assume some subs-criteria in our initial problem. In this assumption, the relationship between criteria and alternative is not symmetric (in our previous work we assumed a symmetric situation).

For example in this new assumption humidity has no effect on the inspection. To get a better understanding of the situation we illustrate the initial problem in Figure 2. In the second stage of DSS, the decision maker translates the linguistic variables to fuzzy sets. To do this we need to have an accurate data mining of the component. As mentioned in our previous work [8], the one-dimensional membership functions have different shapes such as triangular, trapezoidal or Gaussian.

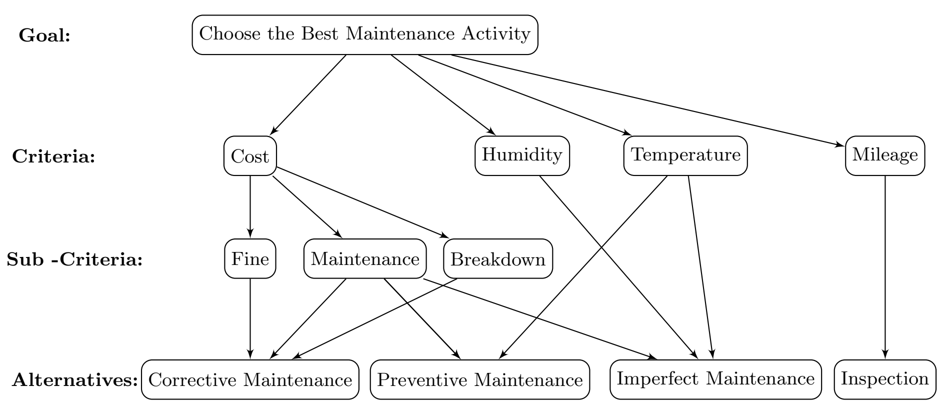


Fig. 2. Problem Statement with four main criteria and three subs-criteria

We use the triangular membership function in this work which is shown in Figure 3.

As Figure 3 indicates we need to use a binary comparison to evaluate the criteria. By a binary comparison we mean two criteria compared with each other.The degree of importance between two criteria can be divided into several groups such as: equal importance, weak, strong, very strong, etc.

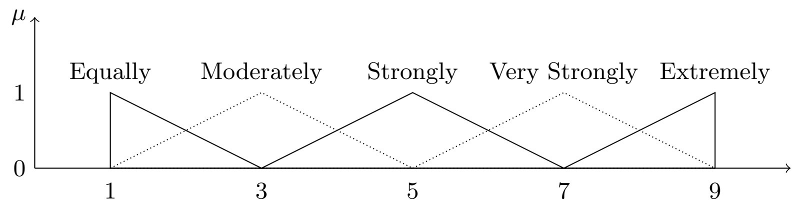


Fig. 3. The five Fuzzy Membership Functions for our linguistic variables

To translate the degree of importance for different criteria to fuzzy numbers we use Table 1.

Fuzzy Number	Description	Triangular fuzzy scale	Domain	Membership function
$\tilde{9}$	Absolute Importance	$(7, 9, 9)$	$7 \le x \le 9$	$(x-7)/(9-7)$
$\tilde{7}$	Demonstrated Importance	(5,7,9)	$7 \le x \le 9$	$(9-x)/(9-7)$
			$5 \le x \le 7$	$(x-5)/(7-5)$
$\tilde{5}$	Strong Importance	(3,5,7)	$5 \le x \le 7$	$(7-x)/(7-5)$
			$3 \le x \le 5$	$(x-3)/(5-3)$
$\tilde{3}$	Weak Importance	(1,3,5)	$3 \le x \le 5$	$(5-x)/(5-3)$
			$1 \le x \le 3$	$(x-1)/(3-1)$
$\tilde{1}$	Equal Importance	(1,1,3)	$1 \le x \le 3$	$(3-x)/(3-1)$
1	Exactly Equal Importance	(1,1,1)		

Table 1. The fuzzy scale of importance

By using Table 1, a decision maker is able to translate the degree of importance for a criterion to a fuzzy number. After performing data mining, study of the history of the component and health status monitoring, we can evaluate how a criterion has effect on the efficiency of the component. For example, by studying the warranty data for the cooling system in different geographic regions, we get an overview of the effect of temperature and humidity on the deterioration rate of the cooling system. In stage three, we use AHP in a fuzzy environment to find the best maintenance activity for the component.

The fuzzy AHP approach is based on the judgment matrix with triangle fuzzy numbers (TFN). Let n be the number of criteria then the $n \times n$ -judgment matrix can be formulate as [9]:

$$\tilde{A} = \begin{pmatrix} (111) & \tilde{a}_{12} & \dots & \tilde{a}_{1n} \\ \tilde{a}_{21} & (111) & \dots & \tilde{a}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{n1} & \tilde{a}_{n2} & \dots & (111) \end{pmatrix} \tag{8}$$

where matrix $\tilde{a}_{ij} = \tilde{a}_{ij}^{-1}$ and $M_{ij} = (l_{ij}, m_{ij}, u_{ij})$.

If we have a group of decision makers, TFN represents three points of views, the first element indicate the minimum value, the second element represents the medium value and the last element is the maximum value [9]. In this step we need to calculate the value of fuzzy synthetic extent $\tilde{S}_i$ for every row in matrix $\tilde{A}$. $\tilde{S}_i$ is a fuzzy number as well and can be found by:

$$\tilde{S}_i = \sum_{j=1}^{m} M_{ij} \otimes \left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij} \right]^{-1}, i = 1, 2, ..., n \tag{9}$$

where $\otimes$ is fuzzy multiplication and $\sum_{j=1}^{n} M_i^j$ is a fuzzy triangular number which can be calculated as:

$$(l_i, m_i, u_i) = \sum_{j=1}^{m} M_{ij} = \left(\sum_{j=1}^{m} l_{ij}, \sum_{j=1}^{m} m_{ij}, \sum_{j=1}^{m} u_{ij} \right),$$

$$\sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij} = \left(\sum_{i=1}^{n} l_{ij}, \sum_{i=1}^{n} m_{ij}, \sum_{i=1}^{n} u_{ij} \right) \tag{10}$$

$$\left[\sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij} \right]^{-1} = \left(\frac{1}{\sum_{i=1}^{n} l_{ij}}, \frac{1}{\sum_{i=1}^{n} m_{ij}}, \frac{1}{\sum_{i=1}^{n} u_{ij}} \right) \tag{11}$$

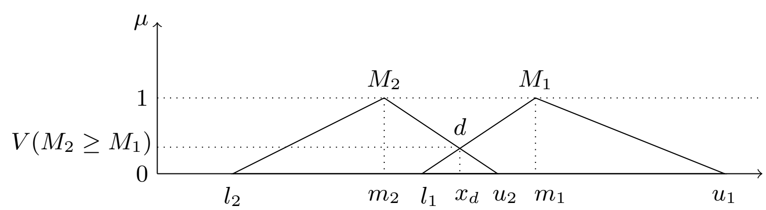


Fig. 4. The degree of possibility for $M_2 \geq M_1$

Let $M_1 = (l_1, m_1, u_1)$ and $M_2 = (l_2, m_2, u_2)$ be triangular fuzzy numbers, according to Figure 4 the degree of possibility of M_1 to M_2 can be defined as [3]:

$$V(M_2 \geq M_1) = \text{hgt}(M_1 \cap M_1) = \mu_{M_2}(d) = \begin{cases} 1 & if\ m_2 \geq m_1 \\ 0 & if\ l_1 \geq u_2 \\ \frac{u_1 - l_2}{(u_1 - l_2) + (m_2 - m_1)} & otherwise \end{cases} \tag{12}$$

Where x_d in Figure 4 indicates the part of highest intersection point d between μ_{M_1} and μ_{M_2} [5]. As last step, we need to measure the weights for the criteria and alternatives, for this we use the following formula:

$$d'(A_i) = \min V(S_i \geq S_k) \qquad k = 1, 2, ..., n, k \neq i$$

the weight vector is also given by[5]:

$$W'(A_i) = (d'(A_1), d'(A_2), ..., d'(A_n))^T \qquad A_i(i = 1, 2, ..., n) \tag{13}$$

To calculate the final weights we need to normalize Equation (13) as follow:

$$W(A_i) = (d(A_1), d(A_2), ..., d(A_n))^T \tag{14}$$

4 A numerical example

To get a better understanding of using AHP technique, we analyze a belt tensioner which is a component in internal combustion engine cooling systems in heavy trucks as a case study. In this section, we try to find the best maintenance activity for this component by using fuzzy AHP. As mentioned before we need to identify the criteria which have direct effect on the system efficiency. Based on our previous work, we use RapidMiner for data mining and estimate the relationship between criteria and alternatives in the fuzzy environment by using Figure 3 and Table 1.

Various maintenance activities can be offered by the companies. To define the alternatives, the decision maker need to study the company's policies, the customer's perspective and requirements, which depend on the company's task operating systems.

According to Figure 2, we define corrective maintenance, preventive maintenance, imperfect maintenance and inspection as the alternatives.

To solve this DSS problem we identify four main criteria: cost, humidity, temperature and mileage and three subs-criteria: fine cost, maintenance cost and breakdown cost. The fuzzy pairwise comparison matrices for the alternatives becomes:

C_1	A_1	A_2	A_3	A_4
A_1	$\tilde{1}$	$\tilde{2}$	$\tilde{3}$	$\tilde{3}$
A_2	$\tilde{2}^{-1}$	$\tilde{1}$	$\tilde{2}$	$\tilde{3}$
A_3	$\tilde{3}^{-1}$	$\tilde{2}^{-1}$	$\tilde{1}$	$\tilde{2}$
A_4	$\tilde{3}^{-1}$	$\tilde{3}^{-1}$	$\tilde{2}^{-1}$	$\tilde{1}$

C_2	A_1	A_2	A_3	A_4
A_1	$\tilde{1}$	$\tilde{9}$	$\tilde{7}$	$\tilde{6}$
A_2	$\tilde{9}^{-1}$	$\tilde{1}$	$\tilde{5}$	$\tilde{3}$
A_3	$\tilde{7}^{-1}$	$\tilde{5}^{-1}$	$\tilde{1}$	$\tilde{2}$
A_4	$\tilde{6}^{-1}$	$\tilde{3}^{-1}$	$\tilde{2}^{-1}$	$\tilde{1}$

C_3	A_1	A_2	A_3	A_4
A_1	$\tilde{1}$	$\tilde{4}$	$\tilde{3}$	$\tilde{1}$
A_2	$\tilde{4}^{-1}$	$\tilde{1}$	$\tilde{3}$	$\tilde{1}$
A_3	$\tilde{3}^{-1}$	$\tilde{3}^{-1}$	$\tilde{1}$	$\tilde{1}$
A_4	$\tilde{1}^{-1}$	$\tilde{1}^{-1}$	$\tilde{1}^{-1}$	$\tilde{1}$

C_4	A_1	A_2	A_3	A_4
A_1	$\tilde{1}$	$\tilde{6}$	$\tilde{9}$	$\tilde{2}$
A_2	$\tilde{6}^{-1}$	$\tilde{1}$	$\tilde{6}$	$\tilde{3}$
A_3	$\tilde{9}^{-1}$	$\tilde{6}^{-1}$	$\tilde{1}$	$\tilde{7}$
A_4	$\tilde{2}^{-1}$	$\tilde{3}^{-1}$	$\tilde{7}^{-1}$	$\tilde{1}$

C_5	A_1	A_2	A_3	A_4
A_1	$\tilde{1}$	$\tilde{8}$	$\tilde{8}$	$\tilde{1}$
A_2	$\tilde{8}^{-1}$	$\tilde{1}$	$\tilde{8}$	$\tilde{2}$
A_3	$\tilde{8}^{-1}$	$\tilde{8}^{-1}$	$\tilde{1}$	$\tilde{3}$
A_4	$\tilde{1}^{-1}$	$\tilde{2}^{-1}$	$\tilde{3}^{-1}$	$\tilde{1}$

C_6	A_1	A_2	A_3	A_4
A_1	$\tilde{1}$	$\tilde{5}$	$\tilde{6}$	$\tilde{9}$
A_2	$\tilde{5}^{-1}$	$\tilde{1}$	$\tilde{4}$	$\tilde{8}$
A_3	$\tilde{6}^{-1}$	$\tilde{4}^{-1}$	$\tilde{1}$	$\tilde{7}$
A_4	$\tilde{9}^{-1}$	$\tilde{8}^{-1}$	$\tilde{7}^{-1}$	$\tilde{1}$

		Cost					
	Fine	Maintenance	Breakdown	Humidity	Temperature	Mileage	$\sum_{j=1}^{n} M_{ij}$
Cost: Fine	(1,1,1)	(5,7,9)	(7,9,11)	(1,3,5)	(1,3,5)	(1,3,5)	(16,26,36)
Cost: Maintenance	$(\frac{1}{9},\frac{1}{7},\frac{1}{5})$	(1,1,1)	(5,7,9)	(1,3,5)	(1,3,5)	(3,5,7)	(11.1,19.1,27.2)
Cost: Breakdown	$(\frac{1}{11},\frac{1}{9},\frac{1}{7})$	$(\frac{1}{9},\frac{1}{7},\frac{1}{5})$	(1,1,1)	(1,3,5)	(1,3,5)	(3,5,7)	(6.2,12.2,18.3)
Humidity	$(\frac{1}{5},\frac{1}{3},1)$	$(\frac{1}{5},\frac{1}{3},1)$	$(\frac{1}{5},\frac{1}{3},1)$	(1,1,1)	(7,9,11)	(5,7,9)	(13.6, 18, 24)
Temperature	$(\frac{1}{5},\frac{1}{3},1)$	$(\frac{1}{5},\frac{1}{3},1)$	$(\frac{1}{5},\frac{1}{3},1)$	$(\frac{1}{11},\frac{1}{9},\frac{1}{7})$	(1,1,1)	(3,5,7)	(4.6,7.1,11.1)
Mileage	$(\frac{1}{5},\frac{1}{3},1)$	$(\frac{1}{7},\frac{1}{5},\frac{1}{3})$	$(\frac{1}{7},\frac{1}{5},\frac{1}{3})$	$(\frac{1}{9},\frac{1}{7},\frac{1}{5})$	$(\frac{1}{7},\frac{1}{5},\frac{1}{3})$	(1,1,1)	(1.7, 2, 3.2)

Table 2. The pairwise comparison matrix for the criteria

Table (2) represents the fuzzy pairwise comparison matrix for the criteria. By using Equations (10) and (11) we get

$$\sum_{i=1}^{n}\sum_{j=1}^{m} M_{ij} = (53.2, 84.4, 119.8) \quad , \quad \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{ij}\right]^{-1} = (0.008, 0.01, 0.01).$$

$$\begin{aligned}
S_1 &= (16, 26, 36) \otimes (0.008, 0.01, 0.01) = (0.128, 0.26, 0.36)\\
S_2 &= (11.1, 19.1, 27.2) \otimes (0.008, 0.01, 0.01) = (0.088, 0.191, 0.272)\\
S_3 &= (6.2, 12.2, 18.3) \otimes (0.008, 0.01, 0.01) = (0.0496, 0.122, 0.183)\\
S_4 &= (13.6, 18, 24) \otimes (0.008, 0.01, 0.01) = (0.108, 0.18, 0.24)\\
S_5 &= (4.6, 7.1, 11.1) \otimes (0.008, 0.01, 0.01) = (0.036, 0.071, 0.111)\\
S_6 &= (1.7, 2, 3.2) \otimes (0.008, 0.01, 0.01) = (0.0136, 0.02, 0.032)
\end{aligned}$$

We determine the weight rating for each criteria by using Equation (12), define $V_{ij} = V(S_i \geq S_j)$ then:

$$\begin{aligned}
&V_{12} = 1, V_{13} = 1, V_{14} = 1, V_{15} = 1, V_{16} = 1\\
&V_{21} = 0.67, V_{23} = 1, V_{24} = 1, V_{25} = 1, V_{26} = 1\\
&V_{31} = 0.28, V_{32} = 0.58, V_{34} = 0.56, V_{35} = 1, V_{36} = 1\\
&V_{41} = 0.58, V_{42} = 0.93, V_{43} = 1, V_{45} = 1, V_{46} = 1\\
&V_{51} = 0.09, V_{52} = 0.16, V_{53} = 0.54, V_{54} = 0.026, V_{56} = 1\\
&V_{61} = 0.66, V_{62} = 0.48, V_{63} = 0.20, V_{64} = 0.90, V_{65} = 0.072
\end{aligned}$$

In this stage we calculate normalized and unnormalized weights for the criteria by using Equations (13) and (14). The results for this step have been summarized in Table 3.

In the next stage we compare the weights for every single criteria with the alternatives, to avoid lengthy calculations we summarize the result in Table 4:

Criteria	Unnormalized Weight	Normalized weight
C_1	1	0.38
C_2	0.67	0.25
C_3	0.28	0.10
C_4	0.58	0.22
C_5	0.026	0.009
C_6	0.072	0.02

Table 3. The weight of the criteria

		A_1	A_2	A_3	A_4	$\sum_{i=1}^{n} M_{ij}$	Unnormalized Weight	Normalized Weight
	A_1	(1,1,1)	(1,2,4)	(1,3,5)	(1,3,5)	(4,9,15)	1	0.333
C_1	A_2	$(\frac{1}{4}, \frac{1}{2}, 1)$	(1,1,1)	(1,2,4)	(1,3,5)	(3.25,6.5,11)	0.882	0.294
	A_3	$(\frac{1}{5}, \frac{1}{3}, 1)$	$(\frac{1}{4}, \frac{1}{2}, 1)$	(1,1,1)	(1,2,4)	(2.45,3.83,7)	0.680	0.226
	A_4	$(\frac{1}{5}, \frac{1}{3}, 1)$	$(\frac{1}{5}, \frac{1}{3}, 1)$	$(\frac{1}{4}, \frac{1}{2}, 1)$	(1,1,1)	(1.65,2.1,4)	0.434	0.144
	A_1	(1,1,1)	(7,9,11)	(5,7,9)	(4,6,8)	(17,23,29)	1	1.036
C_2	A_2	$(\frac{1}{11}, \frac{1}{9}, \frac{1}{7})$	(1,1,1)	(3,5,7)	(1,3,5)	(5,9.1,13.1)	0.338	0.172
	A_3	$(\frac{1}{9}, \frac{1}{7}, \frac{1}{5})$	$(\frac{1}{7}, \frac{1}{5}, \frac{1}{3})$	(1,1,1)	(1,2,4)	(2.2,3.3,5.5)	0.445	0.223
	A_4	$(\frac{1}{8}, \frac{1}{6}, \frac{1}{4})$	$(\frac{1}{5}, \frac{1}{3}, 1)$	(1, 2, 4)	(1,1,1)	(2.3,3.5,6.2)	0.187	0.095
	A_1	(1,1,1)	(2,4,6)	(1,3,5)	(1,1,1)	(5,9,13)	1	0.460
C_3	A_2	$(\frac{1}{6}, \frac{1}{4}, \frac{1}{2})$	(1,1,1)	(1,3,5)	(1,1,1)	(3.1,5.2,7.5)	0.646	0.297
	A_3	$(\frac{1}{5}, \frac{1}{3}, 1)$	$(\frac{1}{5}, \frac{1}{3}, 1)$	(1,1,1)	(1,1,1)	(2.4,2.6,4)	0.239	0.110
	A_4	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)	(1,1,1)	(4,4,4)	0.287	0.132
	A_1	(1,1,1)	(4,6,8)	(7,9,11)	(1,2,4)	(13,18,24)	1	0.520
C_4	A_2	$(\frac{1}{8}, \frac{1}{6}, \frac{1}{4})$	(1,1,1)	(4,6,8)	(1,3,5)	(6.1,10.1,14.2)	0.574	0.298
	A_3	$(\frac{1}{11}, \frac{1}{9}, \frac{1}{7})$	$(\frac{1}{8}, \frac{1}{6}, \frac{1}{4})$	(1,1,1)	(5,7,9)	(6.2,8.2,10.3)	0.344	0.298
	A_4	$(\frac{1}{4}, \frac{1}{2}, 1)$	$(\frac{1}{9}, \frac{1}{7}, 1)$	$(\frac{1}{9}, \frac{1}{7}, \frac{1}{5})$	(1,1,1)	(1.4,1.7,3.2)	0.005	0.002
	A_1	(1,1,1)	(6,8,10)	(6,8,10)	(1,1,1)	(14,18,22)	1	0.595
C_5	A_2	$(\frac{1}{10}, \frac{1}{8}, \frac{1}{6})$	(1,1,1)	(6,8,10)	(1,2,4)	(8.1,11.1,15.1)	0.584	0.347
	A_3	(1, 1, 1)	$(\frac{1}{10}, \frac{1}{8}, \frac{1}{6})$	(1,1,1)	(1,3,5)	(3.1,5.1,7.1)	0.077	0.045
	A_4	(1, 1, 1)	$(\frac{1}{4}, \frac{1}{2}, 1)$	$(\frac{1}{5}, \frac{1}{3}, 1)$	(1,1,1)	(2.45,2.8,4)	0.019	0.011
	A_1	(1,1,1)	(3,5,7)	(4,6,8)	(7,9,11)	(15,21,27)	1	0.409
C_6	A_2	$(\frac{1}{7}, \frac{1}{5}, \frac{1}{3})$	(1,1,1)	(2,4,6)	(6,8,10)	(9.1,13.2,17.3)	0.620	0.253
	A_3	$(\frac{1}{8}, \frac{1}{6}, \frac{1}{4})$	$(\frac{1}{6}, \frac{1}{4}, \frac{1}{2})$	(1,1,1)	(5,7,9)	(6.2,8.4,10.7)	0.214	0.098
	A_4	$(\frac{1}{11}, \frac{1}{9}, \frac{1}{7})$	$(\frac{1}{10}, \frac{1}{8}, \frac{1}{6})$	$(\frac{1}{9}, \frac{1}{7}, \frac{1}{5})$	(1,1,1)	(1.3,1.3,1.5)	0.608	0.248

Table 4. The pairwise comparison matrix between the criteria and alternatives

Table 5 represents the prioritizes for the alternatives, the best alternative is an alternative which has maximum weight:

	C_1	C_2	C_3	C_4	C_5	C_6	
Weight	0.38	0.25	0.10	0.22	0.009	0.02	Sum
A_1	0.333	1.036	0.460	0.520	0.595	0.409	0.559
A_2	0.294	0.172	0.297	0.298	0.347	0.253	0.258
A_3	0.226	0.223	0.110	0.298	0.045	0.098	0.220
A_4	0.144	0.095	0.132	0.002	0.011	0.248	0.097

Table 5. Prioritisation of alternatives

As we see A_1 ranked first which indicates to the corrective maintenance for this component, preventive maintenance and imperfect maintenance are two other best maintenance activities.

5 Planning with Markov Decision Processes (MDP)

When, from the methods presented in the previous sections, we have decided on what actions we want to consider in our maintenance policy, we need to analyze what actions to perform at different system states and times. A natural initial method to this respect is that of optimizing a Markov Decision Process (MDP).

To be able to find the optimal maintenance actions for each point in time (and state) we can model the system by a standard MDP tuple (S, A, R, P), where:

- S is the set of possible states.
- $A(s)$ is the set of possible actions when in the state s.
- $R(s, a, s')$ is the reward acquired from transiting to the state s' if the action a is taken when in state s.
- $P(s, a, s')$ is the probability of transiting to the state s' if the action a is taken when in state s.

In modeling a complex system care must be taken to get enough details to make the analysis relevant, but not so much detail as to make the analysis too cumbersome. A simple model will provide better transparency but a more detailed model will allow for a maintenance policy that actually optimizes the reality better.

Too exemplify a useful variant of this method consider the following model.

We consider 10 time steps and one component subject to maintenance with three stages of deterioration. See Figure 5 for a description of the possible states. Note that we in each possible transition must increase the time step by exactly 1. The deterioration states are numbered $1, 2, 3$ with 1 being the least and 3 being the most deteriorated state.

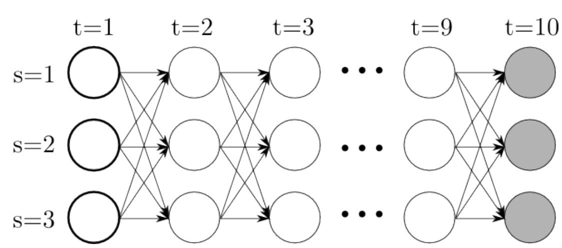


Fig. 5. The 30 possible states indexed by s, t where s is the deterioration state and t is the time step. The possible initial states are at $t = 1$ and the terminating states are at $t = 10$. Only transitions that increases the time step by one are allowed.

We have three different actions, all allowed at each state, $A(s) = \{1, 2, 3\}$, where the action 1 is the do nothing (normal operation) action, 2 is an imperfect maintenance not necessarily restoring the component to deterioration state 1, and 3 is perfect maintenance always restoring the component to deterioration state 1. The reward function $R(s, a, s')$ and the probability function $P(s, a, s')$

can now be described by 3×3 matrices R_1, R_2, R_3 and P_1, P_2, P_3 with components R_{aij} and P_{aij} defining the reward/probability going from deterioration state i to j through the action a.

$$R_1 = \begin{pmatrix} r_0(t) & r_0(t) & r_0(t) \\ r_1 & r_0(t) & r_0(t) \\ r_1 & r_1 & r_0(t) \end{pmatrix} \; R_2 = \begin{pmatrix} r_2 & 0 & 0 \\ r_2 & r_2 & 0 \\ 0 & r_2 & r_2 \end{pmatrix} \; R_3 = \begin{pmatrix} r_3 & 0 & 0 \\ r_3 & 0 & 0 \\ r_3 & 0 & 0 \end{pmatrix}$$

$$P_1 = \begin{pmatrix} p_{11} & 2(1-p_{11})/3 & (1-p_{11})/3 \\ (1-p_{22})/2 & p_{22} & (1-p_{22})/2 \\ 1-p_{33} & 0 & p_{33} \end{pmatrix} \; P_2 = \begin{pmatrix} 1 & 0 & 0 \\ p_u & p_e & 0 \\ 0 & p_u & p_e \end{pmatrix} \; P_3 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

The R_1 matrix defining normal operation has a time dependent reward $r_0(t) > 0$ when the next deterioration is equal or worse (no failure) and a constant reward $r_1 < 0$ when the deterioration state is improved due to corrective maintenance. In R_2 there is a constant reward $r_2 < 0$ when imperfect maintenance is performed, improving the deterioration by one step or having negligible effect. In R_3 the constant reward $r_3 < 0$ is acquired when doing perfect maintenance restoring the deterioration to the best state. The constant probabilities p_{11}, p_{22}, p_{33} define the probabilities of remaining in the same deterioration state when doing action 1. The other elements of P_1 models the normal operation deterioration probabilities, and the failure probabilities resulting in corrective maintenance. The probabilities p_u and p_e are the probabilities of improving the state by one or making no change at all, this models the nature of imperfect maintenance. The P_3 matrix describes a probability of exactly 1 to restore the deterioration to the best state when doing perfect maintenance. Note that P_1, P_2, P_3 are stochastic.

$r_0(t)$	r_1	r_2	r_3	p_{11}	p_{22}	p_{33}	p_u	p_e
$100, t \neq 5$ $1000, t = 5$	-200	-10	-50	0.9	0.8	0.7	0.8	0.2

Fig. 6. Example system variables to illustrate method properties. These values completes the definition of our system model.

This system is now described by a directed acyclic graph, but we have at this point no way of telling which rewards and transitions that should be active at each state, this depends on the actions taken. Using dynamic programming with backward recursion from the terminating states $t = 10$ we are able to for each of the 30 possible states decide upon the optimal action, actually 27 states since we cannot transit from the 3 terminating states. Backward recursion is performed by the recursion:

$$Value(s) = \max_a \sum_{s'} P(s, a, s') \left(R(s, a, s') + Value(s') \right), \tag{15}$$

and we simply collect the actions a from these maximizations for each system state. The resulting state to action function is in the language of MDP denoted a *policy*. Note that the “time” dependence is not present in this recursion, but it is implicit in our choice of possible states. The results we got for this system is:

$$Policy = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 3 & 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 1 & 3 & 3 & 3 & 1 \end{pmatrix}$$

$$Value = \begin{pmatrix} 1699 & 1612 & 1524 & 1436 & 1365 & 376 & 286 & 195 & 100 \\ 1577 & 1490 & 1402 & 1315 & 1162 & 276 & 204 & 137 & 70 \\ 1562 & 1474 & 1386 & 1315 & 918 & 236 & 145 & 50 & 10 \end{pmatrix}$$

It can be seen that if we are in the state of lowest deterioration, row 1, it is optimal to perform a normal operation with no preventive maintenance, there are only ones in the first row of the $Policy$ matrix. When in the state of highest deterioration, row 3, it is usually optimal to perform perfect maintenance, there are mostly 3:s in the last row of the $Policy$ matrix. Exception to this rule is for $t = 5$ where it is optimal to, even if in the worst deterioration state, perform normal operation. The reason for this is the high success reward of $s_0(5) = 1000$ under action 1 at time 5. We can also see that up until $t = 4$ we should perform imperfect maintenance (action 2) if in deterioration state 2 (row 2), but at time $t = 4$ we should perform perfect maintenance from state 2. This is from the backward recursion algorithm anticipating the high reward from normal operation at $t = 5$ and sacrificing a high preventive maintenance cost at $t = 4$ and the subsequent loss of normal operation income of 100 to increase the probability of acquiring the reward of 1000 at $t = 5$. This high reward can be seen in the $Value$ matrix as the jump in total acquired reward from $t = 5$ to column $t = 6$.

This exemplifies the method of MDP in maintenance planning. Note that in this model we could easily have time-dependent R_a and P_a for all transitions, this would not change the ease of the backward recursion optimization, the Markov property is still satisfied. In reality we could have thousands of possible deterioration states and time steps and a more complex transition into future states, with time-steps larger than one, and several different components in the same system. This probabilities and rewards in these complex cases needs to be adapted to historical data and theoretical component models for good model consistency with reality. Note also that the Policy described in the $Policy$ matrix denotes the actions taken for each state when we actually are in that state, but for some organizations this Just In Time policy might not be viable and a plan for future maintenance activities might be desirable. In this case further adaptation through the next step in the decision support framework are required.

6 Conclusions

In this paper, we introduced an approach for strategic maintenance planning. The approach uses a fuzzy AHP technique in order to evaluate different maintenance activities for the various component of cooling system. By using decision making methods in fuzzy environment, we are able to provide several interesting potentials for future extensions. As showed, triangular fuzzy numbers (TFN) can also be defined as a set. In the present paper we considered a set for each fuzzy number (see Table 1). In our previous work [8] we were only interested in the best alternative as the output of the decision making process. The relationship between criteria and alternatives was symmetric and no sub-criteria have been assumed. However, the approach has the capability to propose the second or third best alternatives, and the worst case as well, which is an advantage of using TOPSIS over AHP [7]. We have also shown an example of maintenance planning optimization using MDP (Markov Decision Processes) by a simple but hopefully instructive example. The methods using dynamic programming on MDP optimization problem are simple and efficient in many cases and a good model for optimization of maintenance systems where the Markov property is satisfied.

References

1. N.H. Mateou A.S. Andreou and G.A. Zombanakis. Evolutionary fuzzy cognitive maps: A hybrid system for crisis management and political decision-making. pages 732–743.
2. K. Atanassov. Intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, pages 87–96, 1986.
3. Da-Yong Chang. Applications of the extent analysis method on fuzzy {AHP}. *European Journal of Operational Research*, 95(3):649 – 655, 1996.
4. J. Fan and W. Xie. Subsethood measure: new definitions. fuzzy sets and systems. *Fuzzy Sets and Systems*, pages 201–209, 1999.
5. M. Ahsan Akhtar Hasin Golam Kabir. Evaluation of customer oriented success factors in mobile commerce using fuzzy ahp. *Engineering and Management*, pages 361–386, 2011.
6. P.Dellolme M.Caramia. Multi-objective managment in freight logisticsincreasing capasity, service level and safety with optimization algorithms. pages 12–14, 2008.
7. Mehrdad Saadatmand and Sahar Tahvili. A fuzzy decision support approach for model-based tradeoff analysis of non-functional requirements. In *12th International Conference on Information Technology : New Generations*, April 2015.
8. Sahar Tahvili, Jonas Österberg, Sergei Silvestrov, and Jonas Biteus. Solving complex maintenance planning optimization problems using stochastic simulation and multi-criteria fuzzy decision making. *AIP Conference Proceedings*, 1637, 2014.
9. T. Terano. *Fuzzy Engineering Toward Human Friendly Systems*. Number v. 2. Ohmsha, 1992.
10. Ying-Ming Wang and Kwai-Sang Chin. Fuzzy analytic hierarchy process: A logarithmic fuzzy preference programming methodology. *International Journal of Approximate Reasoning*, 52(4):541 – 553, 2011.
11. J. Yang and J. Watada. Fuzzy clustering analysis of data mining: Application to an accident mining system. *International Journal of Innovative Computing, Information and Control*, page 5715–5724, 2012.
12. Y.Shi. *A Deep Study of Fuzzy Implications*. 2009.
13. C. YU and Y. Luo. A fuzzy optimization method for multi-criteria decision making problem based on the inclusion degrees of intuitionistic fuzzy sets. *Journal of Information and Computing Science*, pages 146–152, 2008.
14. LA. Zadeh. Fuzzy sets. *Information and Control*, page 338–353, 1965.

A Model Selection Method for Heavy-tailed Clustering

Dvora Toledano-Kitai[1], Renata Avros[2], and Zeev Volkovich[3]

[1] Software Engineering Department, ORT Braude College P.O. Box: 21982 Karmiel, ISRAEL
(E-mail: dvora@braude.ac.il)
[2] Software Engineering Department, ORT Braude College P.O. Box: 21982 Karmiel, ISRAEL
(E-mail: ravros@braude.ac.il)
[3] Software Engineering Department, ORT Braude College P.O. Box: 21982 Karmiel, ISRAEL (E-mail: vlvolkov@braude.ac.il)

Abstract. The multivariate Gaussian distributions can be applied to approximate distributional forms arising in clustering. However, the tails of the used Gaussian distributions often do not reflect the real data structure as being sufficiently shorter than actual ones. In this paper we propose an approach to evaluate the components quantity in clustering based on the multivariate Laplace elliptical distributions which are compared by relying upon the Geodesic distance. Possible numbers of clusters are evaluated from the stability standpoint using the scaling parameter of the corresponding Spectral clustering procedure. Clusters quantity exhibiting the most stable behavior of this parameter is accepted as an estimation of the number of clusters in the data. A possible application of the proposed methodology to image segmentation is discussed.
Keywords: Model Selection, Clustering, Laplacian distribution.

1 Introduction

Mixtures of the multivariate Gaussian distribution can be applied to approximate distributional forms arising in clustering. However, the tails of the used Gaussian distributions often do not reflect the real data structure as being suggestively shorter than actual ones. For example, such distributions are arising in image segmentation problems where the object textures are distinguished. Here, the distributions of the appropriate wavelet coefficients are well approximated by the heavy tailed distributions (see, e.g. [6]).

The goal of clustering is to divide a given dataset into dissimilar sub-groups, where resemblance between items belonging to the same group is much higher than between items belonging to other groups. The cluster validation methods are known as one of the systematic activities within the field clustering. These methods are designed to verify the efficacy of a clustering process by defining the true number of clusters in the dataset. The general problem called cluster model selection involves evaluating the possible clustering solutions according to predefined rules and determining the best number of clusters as the one which yields the best clustering result.

16th ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece

This famous problem is a critical challenge in the cluster analysis area which has not yet been appropriately addressed. While, many approaches were proposed to overcome this concern, until now none stand out as the most effective for the general-purpose model selection problem (see, a review of cluster validation methods in [8]).

Applying the model selection for estimation of the true numbers of textures is very crucial in these kinds of tasks. Actually, the selection influences the resolution of the obtained image segmentation.

In this paper we propose an approach to evaluate the components quantity in clustering based on the multivariate Laplace elliptical distributions which are compared by relying upon the Geodesic distance. Possible numbers of clusters are evaluated from the stability standpoint using the scaling parameter of the corresponding Spectral clustering procedure. Clusters quantity exhibiting the most stable behavior of this parameter is accepted as an estimation of the number of clusters in the data. A possible application of the proposed methodology to image segmentation is discussed.

2 The Geodesic Distance for *GGD*

A wide class of Elliptical distributions (Generalized Gaussian Distribution (GGD)) can be defined in terms of their density functions

$$f\left(x|\Sigma,\beta\right)=\frac{\Gamma\left(\frac{m}{2}\right)}{\pi^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)2^{\frac{m}{2\beta}}}\frac{\beta}{|\Sigma|}\cdot\exp\left(-\frac{1}{2}\left(x^{T}\Sigma^{-1}x\right)^{\beta}\right), \tag{1}$$

where

- Σ - is the covariance matrix;
- β - is the shape parameter. $\beta=0.5$-Laplacian distribution and $\beta=1$ -Gaussian distribution;
- m - is the probability space dimension. in case of color images $m=3$.

One of the elliptical possible distributions choices is the Laplacian distribution which will be discussed in this paper. We recognize each GGD as a point on a smooth Riemann manifold. The geodesic manifold distance (GD) provides a powerful tool for calculating the degree of similarity between two GGD distributions (see, [1]):

$$GD\left(\beta,\Sigma_{1}|\beta,\Sigma_{2}\right)=\left[\left(3b_{m}-\frac{1}{4}\right)\sum_{i}\left(r_{2}^{i}\right)^{2}+2\left(b_{m}-\frac{1}{4}\right)\sum_{i<j}r_{2}^{i}\cdot r_{2}^{j}\right]^{\frac{1}{2}}, \tag{2}$$

where

- $r_{2}^{i}\equiv\ln\left(\lambda_{2}^{i}\right)$ and $\lambda_{2}^{i}, i=1,...,m$ are eigenvalues of $\Sigma_{1}^{-1}\Sigma_{2}$
- $b_{m}=\frac{m+1}{4(m+2)}$ - for Laplacian distribution.

2.1 Spectral Clustering

Spectral clustering skills commonly employ the spectrum of a given similarity matrix in order to perform dimensionality reduction for clustering in fewer dimensions. Recall, that this matrix is usually given as $e^{-\frac{d(x_i,x_j)}{2\sigma}}$ where $d(x_i, x_j)$ is a distance between items x_i, x_j, and σ is the scaling parameter. There is a large family of clustering algorithms based on the spectral clustering methodology (see, for example [4], [3], [2]).

A method for model estimation for spectral clustering has been proposed in [7]. The main idea of this approach consists of searching for the number of clusters, for which a statistical estimation of the scaling parameter σ demonstrates the most stable behavior. The algorithm can be presented as such:

Algorithm 1 ***Self-Learning Spectral Clustering*** (X, k, F)
Input

- X *- the data to be clustered;*
- k *- number of clusters;*
- F *- cluster quality function to be minimized.*

Output

- σ^**- an optimal value of the scaling parameter;*
- $\Pi_{k,\ \sigma^*}(X)$*−a partition of* X *into* k *clusters corresponding to* σ^**.*
  ====================

Return

$$\sigma^* = \arg\min_{\sigma}(F(\Pi_{k,\ \sigma}(X) =$$
$$= Spectral\ Clustering(X{,}k{,}\sigma))).$$

As a spectral clustering algorithm the NJW algorithm [5] was applied.

The concept to use the behavior of σ in order to estimate the true number of clusters has been proposed in [7] in the following way: We draw samples from the population and learn values of σ for different number of clusters using partition evaluation criteria. The number of clusters for which the distribution of σ is the most concentrated is accepted as the true one. (Note later that in problems regarding partitions of images, the resulted concentration is at the origin).

In this paper we use a partition quality evaluation function found in the framework of the distance learning methodology. Let us presume that the degree of similarity between pairs of elements of data collection is known:

$$S:\ \{(x_i, x_j)\,;\ if\ x_i\ and\ x_j\ are\ similar$$
$$(belong\ to\ the\ same\ cluster)$$

and

$$D:\ \{(x_i, x_j)\,;\ if\ x_i\ and\ x_j\ are\ not\ similar$$
$$(belong\ to\ different\ clusters).$$

The goal is to learn a distance metric $d(x,y)$ such that all "similar" data points are kept in the same cluster, (i.e., close to each other) while distinguishing the "dissimilar" data points. To this end, we define a distance metric in the form:

$$d_C^2(x,y) = \|x-y\|_A^2 = (x-y)^T \cdot C \cdot (x-y)\,,$$

where C is a positive semi-definite matrix, $C \succ 0$ which is learned. We can formulate a constrained optimization problem where we aim to minimize the sum of similar distances concerning pairs in S while maximizing the sum of dissimilar distances related to pairs in D in the following way:

$$\min_C \sum_{(x_i,x_j)\in S} \|x_i - x_j\|_C^2$$

s.t.

$$\sum_{(x_i,x_j)\in D} \|x_i - x_j\|_C^2 \geq 1,\ C \succ 0.$$

If we suppose that the metric matrix is diagonal then minimizing the function is equivalent to solving the stated optimization problem [9] up to a multiplication of C by a positive constant.

In the spirit of Fisher's linear discriminant analysis we consider the following quality evaluation function:

$$F(\Pi_{k,\ \sigma}(X)) = \frac{\sum_{(x_i,x_j)\in S, i\neq j} \|y_i - y_j\|^2}{\sum_{(x_i,x_j)\in D} \|y_i - y_j\|^2}. \tag{3}$$

3 The General Model

Given a set of items recognized as *GGD* distributions of the same type, we would like to divide this set automatically into subsets resting upon the inner similarity, while choosing the appropriate number of subsets.

A meta algorithm proposed for this purpose is described as follows:

1. Calculate the geodesic distances $d_{i,j}$, between all the items;
2. Randomly draw a symmetric sub-matrices $\mathbb{S}_t$, $t = 1, 2, .., T$ from the distances matrix $\mathbb{D} = (d_{i,j})_{i,j}$;
3. For each entry of number of clusters k and for each sample $\mathbb{S}_t$: apply the Self-Learning Clustering algorithm to find the optimal parameter $\sigma_k(\mathbb{S}_t)$;
4. For each entry of k and the set of samples $\mathbb{S}_t$, $t = 1, 2, ..., T$ construct the distribution of $\{\sigma_k(\mathbb{S}_t)\}$, $t = 1, 2, ..., T$.
5. Determine $k^\star$ the optimal value of k as the one for which the distribution of $\{\sigma_{k^\star}(\mathbb{S}_t)\}$, $t = 1, 2, ..., T$ is the most concentrated at the origin.
6. Execute the Spectral Clustering algorithm on $\mathbb{D}$ with the found true number of clusters $k^\star$ via the optimal parameter $\sigma^\star = \sigma_{k^\star}(\mathbb{D})$ to obtain the partition $\boldsymbol{\Pi}_{k^\star,\sigma^\star}(\mathbb{D})$.

4 Application for image segmentation

As was mentioned above in Section 2, the distributions of the wavelet coefficients calculated for an image parts can be well approximated by means of the *GGD* model, especially via the Laplacian distributions. So different distributions, which can be distinguished by the correlation matrices correspond to different textures.

Wavelet transform is one of the common tools used for analysis and classification of textures in images. This can be attained firstly by modeling the wavelet coefficients while applying statistical analysis of their distributions. A suitable distance for estimating the similarities of these distributions needed to be used to classify textures in the image.

We assume that there is a stable image segmentation provided by the different image textures. In order to discover such a segmentation we apply the methodology described in the meta-algorithm presented in Section 3.

Namely, to identify the number of textures of a given image, we draw T times a symmetric sub-matrix $\mathbb{S}_t$ from the distances matrix $\mathbb{D} = \{d_{i,j}\}$. Then, for every tested number of clusters k, we find the optimal parameter $\sigma_k(\mathbb{S}_t)$ by executing the Self-Learning algorithm on $\mathbb{S}_t$. The true number of textures is chosen as the one for which the distributions of the obtained optimal values $\{\sigma_k(\mathbb{S}_t)\}$, $t = 1, 2, ..., T$ is the most concentrated at 0. Finlay, we apply the Self-Learning Clustering algorithm on $\mathbb{D}$ with the attained true numbers of the textures to reveal the image different textures.

A new automatic image segmentation algorithm can be summarized as follows:

Input

- $\mathbb{X}$ - the image to be clustered;
- K - the maximal tested number of clusters;
- T - the number of drawn sub-matrices;
- p - the sample rate;
- h - the rectangular size ;

1. Divide the image into rectangular sub-images with length and width of size h;
2. In each sub-window: calculate the characterizing wavelet coefficients *GGD* distributions ;
3. Calculate the geodesic distances $d_{i,j}$, between the found distributions.
4. For each of the tested number of clusters $2 \leq k \leq K$:
5. For each $1 \leq t \leq T$:
6. Randomly draw from the distances matrix $\mathbb{D} = (d_{i,j})$, a symmetric sub-matrix $\mathbb{S}_t$ according to the given sample rate p ;
7. Execute the Self-Learning Spectral clustering algorithm on $\mathbb{S}_t$ to obtain the optimal parameter $\sigma_k(\mathbb{S}_t)$;
8. Construct the distribution of $\{\sigma_k(\mathbb{S}_t)\}$, $t = 1, 2, ..., T$;
9. Determine $k^\star$- the optimal value of k as the one for which the distribution of $\{\sigma_{k^\star}(\mathbb{S}_t)\}$, $t = 1, 2, ..., T$ is the most concentrated at the origin.

10. Execute the Spectral Clustering algorithm on $\mathbb{D}$ with the found true number of clusters $k^\star$ via the optimal $\sigma^\star = \sigma_{k^\star}(\mathbb{D})$ to obtain the partition $\boldsymbol{\Pi}_{k^\star,\sigma^\star}(\mathbb{D})$;

As a result of the process, image areas with different textures can be revealed.

5 Numerical Experiments

We provide several experiments in order to demonstrate the capability of the proposed method on color images with diverse natural textures. Each image is transformed into a gray-level using the standard rgb2gray MATLAB procedure. Later, the gray-level image is divided into smaller non-overlapping rectangular windows of size nXm. The size of each window must be sufficiently large (e.g. $n, m > 32$), aiming to get representative histograms of the wavelet coefficients. The discrete Haar wavelet transform is applied with two levels to each one of the windows. Then, for each window, the wavelet coefficients of each subband are grouped, and a Laplacian GGD approximates their distributions.

The last step is performed in a different manner in comparison with [5]. In this article the covariance matrices were calculated between subband coefficients of the RGB components. Here we are dealing only with the most intensive, therefore the covariance matrix is estimated as an auto-covariance one with the lag equaling to 3.

The geodesic pairwise distances matrix is figured for all the windows and the similarity matrix S is used for the spectral clustering. The parameters values are as such:

- $K = 5$;
- $T = 50$;
- $p = 10$;
- h - each image is divided into 400 small windows.

In order to characterize the concentration at the origin, the area of the scale parameter values obtained for each number of clusters was divided into 10 equal sized parts. The frequency of the lowest interval is taken as the concentration attitude. Clearly, big values of this parameter can approve high concentration at the origin. Each experiment is repeated 10 times, and the results are presented via error bar plots of the mean of this indicator calculated for the quality cluster function (3).

The first image considered is presented in Fig. 1.

As can be seen in Fig. 2, the proposed algorithm managed to detect the correct number of clusters in the image corresponding to three different image textures.

In the second case, a wall image with different surfaces is chosen (Fig. 3).

Fig. 4 shows that two different textures are exactly recognized in spite of the color difference within the texture segments.

It can be seen in Fig. 6, that although four different textures are not clearly separated inside the image, the true number of clusters has been, clearly detected.

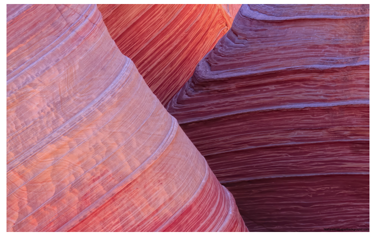


Fig. 1. The first segmented image

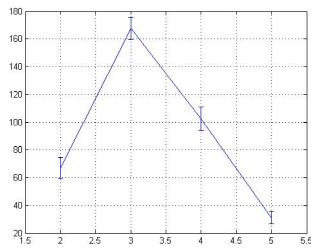


Fig. 2. Error bar plot of first segmented image.

Fig. 3. The second segmented image.

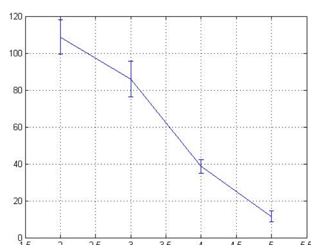


Fig. 4. Error bar plot of the frequency of the lowest interval constructed for the second image.

Fig. 5. The third segmented image.

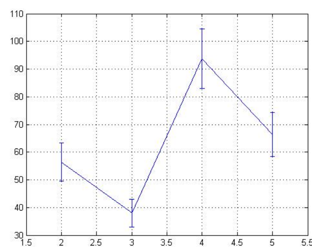


Fig. 6. Error bar plot of the frequency of the lowest interval constructed for the third image.

6 Summary and Conclusions

We proposed a new method intended to model section in the case where items of the clustered dataset is associated with the multivariate Laplacian distribution. Partitions constructed for several clusters quantities are compared by means of the behavior of the sample scale parameter using the consequent Spectral clustering procedure. The most stable behavior of this parameter provides an estimation of the number of clusters in the data. Numerical experiments delivered on several real images exhibits a high ability of the method to reveal correct segmentations based on the object's inner textures.

References

1. M. Berkane, K. Oden, and P.M. Bentler, Geodesic Estimation in Elliptical Distributions, *Journal of Multivariate Analysis*, 63, no. MV971690, 35-46, 1997.
2. M. Filippone, F. Camastra, F. Masulli and S. Rovetta, A survey of kernel and spectral methods for clustering, *Pattern Recognition*, 41 (1), 176-190, 2008.
3. U. V. Luxburg, A tutorial on spectral clustering, *Statistics and Computing*, 17 (4), 395-416, 2007.
4. M. Nascimento and A. D. Carvalho, Spectral methods for graph clustering a survey, *European Journal Of Operational Research 21*, 16 (2), 221-231, 2011.
5. A. Y. Ng, M. I. Jordan and Y. Weiss, On spectral clustering: analysis and an algorithm, *In Advances in Neural Information Processing Systems 14 (NIPS 2001)*, 849-856, 2001.

6. G. Verdoolaege and P. Scheunders, Geodesics on the Manifold of Multivariate Generalized Gaussian Distributions with an Application to Multicomponent Texture Discrimination, *International Journal of Computer Vision*, 95 (3), 265-286, 2011
7. Z. Volkovich and R. Avros, Model Selection and Stability in Spectral Clustering, *KDIR 2012 - Proceedings of the International Conference on Knowledge*, Barcelona, Spain, October 2012.
8. Z. Volkovich, Z. Barzily, G.-W. Weber, D. Toledano-Kitai and R. Avros, Resampling Approach for Cluster Model Selection, *Machine Learning*, 85 (1-2), 209-248, 2011.
9. E. P. Xing, A. Y. Ng, M. I. Jordan and S. Russell, Distance metric learning, with application to clustering with side-information. *Advances in Neural In-formation Processing Systems*, 15 (NIPS 2002), 505-512, 2002.

Entropy and Information Extensions through the Generalized Normal Distribution

Thomas L. Toulias[1], and Christos P. Kitsos[2]

[1] Technological Educational Institute of Athens (until Mar. 2014), 12243 Egaleo, Athens, Greece
(E-mail: ttoulias@ebox.gr)

[2] Department of Informatics, Technological Educational Institute of Athens, 12243 Egaleo, Athens, Greece
(E-mail: xkitsos@teiath.gr)

Abstract. This paper presents and discusses two generalized forms of the Shannon entropy, as well as a generalized information measure. These measures are applied on the γ–order Normal distribution $\mathcal{N}_\gamma$, which is linked to the generalized Fisher's entropy type information measure. The Kullback–Leibler divergence between two $\mathcal{N}_\gamma$ distributions is also discussed. The γ–order Lognormal distribution $\mathcal{LN}_\gamma$, the left/right truncated cases of $\mathcal{N}_\gamma$ and $\mathcal{LN}_\gamma$ distributions, and a two–way asymmetric form of the $\mathcal{N}_\gamma$ distribution are also discussed under the light of the γ–order generalized Normal distribution.

Keywords: Fisher's entropy type information measure, γ–order Normal distribution, Rényi entropy, entropy type information measures.

1 Introduction

There is a number of well known of information measures which all attempt to quantify the uncertainty of an outcome of an experiment (industrial, biological, physical etc), and therefore we are referring to information as a measure of uncertainty. Signal Processing and Cryptography are two main fields of applicable aspects of information measures, see Bauer [1], Hoffstein *et al.* [7] and Stinson [17].

In principle, the information measures are divided into three main lines of thought: parametric (Fisher's information being a typical example), non parametric (with Shannon information measure being the most well known) and entropy type, see Cover and Thomas [4], which are adopted in this paper.

Let X be a multivariate random variable (r.v.) with parameter vector $\theta = (\theta_1, \theta_2, \dots, \theta_p)^{\mathrm{T}} \in \mathbb{R}^{p\times 1}$ and p.d.f. $f_X = f_X(x;\theta)$, $x \in \mathbb{R}^{p\times 1}$. The parametric type Fisher's information measure $\mathrm{I_F}(X;\theta)$ (also denoted as $\mathrm{I}_\theta(X)$) defined as the covariance of $\nabla_\theta \log f_X(X;\theta)$ (where ∇_θ is the gradient with respect to the parameters θ_i, $i = 1, 2, \dots, p$) is a parametric type information measure, expressed also as

$$\mathrm{I}_\theta(X) = \mathrm{Cov}\left(\nabla_\theta \log f_X(X;\theta)\right) = \mathrm{E}_\theta\left[\|\nabla_\theta \log f_X(X)\|^2\right],$$

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*

where $\|\cdot\|$ is the usual $\mathscr{L}^2(\mathbb{R}^p)$ norm, while $\mathrm{E}_\theta[\cdot]$ denotes the expected value operator applied to random variables, with respect to parameter θ.

The Fisher's entropy type information measure $\mathrm{J}(X)$, of a r.v. X with p.d.f. f on $\mathbb{R}^p$ is defined, respectively, as the covariance of the r.v. $\nabla \log f(X)$, i.e. $\mathrm{J}(X) := \mathrm{E}[\|\nabla \log f(X)\|^2]$, with $\mathrm{E}[\cdot]$ denoting the usual expected value operator of a random variable with respect to the its probability density function (p.d.f.) Hence, $\mathrm{J}(X)$ equals

$$\mathrm{J}(X) = \int\limits_{\mathbb{R}^p} f(x)\|\nabla \log f(x)\|^2 dx = \int\limits_{\mathbb{R}^p} \tfrac{1}{f(x)}\|\nabla f(x)\|^2 dx = 4\int\limits_{\mathbb{R}^p} \left\|\nabla\sqrt{f(x)}\right\|^2 dx. \tag{1}$$

See also [11] for some extensions.

Consider now the Vajda's parametric type measure of information $\mathrm{I_V}(X;\,\theta,\alpha)$, which is a generalization of the Fisher's (parametric) information, $\mathrm{I}_\theta(X)$, as defined by Vajda [20],

$$\mathrm{I}_V(X;\,\theta,\alpha) := \mathrm{E}_\theta[\|\nabla_\theta \log f(X)\|^\alpha], \quad \alpha \geq 1. \tag{2}$$

Similarly, the entropy type version of (2) generalizes the Fisher's entropy type information $\mathrm{J}(X)$, and is defined as

$$\mathrm{J}_\alpha(X) := \mathrm{E}[\|\nabla \log f(X)\|^\alpha], \quad \alpha \geq 1, \tag{3}$$

see Kitsos and Tavoularis [9,10]. We shall refer to $\mathrm{J}_\alpha(X)$ as the generalized Fisher's entropy type information measure or α–GFI. The second–GFI is reduced to the usual J, i.e. $\mathrm{J}_2(X) = \mathrm{J}(X)$. Equivalently, from the definition of the α–GFI above we can obtain

$$\mathrm{J}_\alpha(X) = \int\limits_{\mathbb{R}^p} \|\nabla \log f(x)\|^\alpha f(x)dx = \int\limits_{\mathbb{R}^p} \|\nabla f(x)\|^\alpha f^{1-\alpha}(x)dx = \alpha^\alpha \int\limits_{\mathbb{R}^p} \|\nabla f^{1/\alpha}(x)\|^\alpha dx. \tag{4}$$

The Shannon (or relative) entropy $\mathrm{H}(X)$ of a continuous r.v. X with p.d.f. f is a also a fundamental concept in Information Theory, applied heavily in Cryptography, Stinson [17], and defined as,

$$\mathrm{H}(X) := \mathrm{E}[\log f(X)] = \int\limits_{\mathbb{R}^p} f(x)\log f(x)dx, \tag{5}$$

Cover and Thomas [4], where we omit here the usual minus sign.

The corresponding entropy power $\mathrm{N}(X)$ is defined by

$$\mathrm{N}(X) := \nu e^{\frac{2}{p}\mathrm{H}(X)}, \tag{6}$$

where $\nu := (2\pi e)^{-1}$. The generalized entropy power $\mathrm{N}_\alpha(X)$, introduced in Kitsos and Tavoularis [9], is of the form

$$\mathrm{N}_\alpha(X) := \nu_\alpha e^{\frac{\alpha}{p}\mathrm{H}(X)}, \tag{7}$$

with normalizing factor ν_α given by the appropriate generalization of ν, namely

$$\nu_\alpha := \left(\tfrac{\alpha-1}{\alpha e}\right)^{\alpha-1} \pi^{-\frac{\alpha}{2}} \left[\frac{\Gamma\left(\frac{p}{2}+1\right)}{\Gamma\left(p\frac{\alpha-1}{\alpha}+1\right)}\right]^{\frac{\alpha}{p}}, \quad \alpha \in \mathbb{R} \setminus [0,1]. \tag{8}$$

For the parameter case of $\alpha = 2$, (7) is reduced to the known entropy power $\mathrm{N}(X)$, i.e. $\mathrm{N}_2(X) = \mathrm{N}(X)$ and $\nu_2 = \nu$.

The known information inequality $\mathrm{J}(X)\mathrm{N}(X) \geq p$ still holds under the generalized entropy type Fisher's information, as $\mathrm{J}_\alpha(X)\mathrm{N}_\alpha(X) \geq p$, $\alpha > 1$, see Kitsos and Tavoularis [9], who also worked on a generalization of the Gramer–Rao inequality.

Through the generalized entropy power N_α a generalized form of the usual Shannon entropy can be produced, as the Shannon entropy whose entropy power is N_α (instead of the usual N), i.e.

$$\mathrm{N}_\alpha(X) := \nu \exp\{\tfrac{2}{p}\mathrm{H}_\alpha(X)\}, \quad \alpha \in \mathbb{R} \setminus [0,\, 1], \tag{9}$$

called the *generalized Shannon entropy*, or *α–Shannon entropy*. Therefore, from (7) a linear relation between the generalized Shannon entropy $\mathrm{H}_\alpha(X)$ and the usual Shannon entropy $\mathrm{H}(X)$ is obtained, i.e.

$$\mathrm{H}_\alpha(X) = \tfrac{p}{2}\log\tfrac{\nu_\alpha}{\nu} + \tfrac{\alpha}{2}\mathrm{H}(X), \quad \alpha \in \mathbb{R} \setminus [0,\, 1]. \tag{10}$$

Essentially, (10) represents a linear transformation of $\mathrm{H}(X)$, which depends on the parameter α and the dimension $p \in \mathbb{N}$. It is also clear that the generalized Shannon entropy for $\alpha = 2$ is the usual Shannon entropy, i.e. $\mathrm{H}_2 = \mathrm{H}$.

2 Entropy and information of the generalized Normal distribution

Kitsos and Tavoularis in [9,10] introduced and studied a three parameter generalization of the multivariate Normal distribution, and its construction is related to the generalized entropy power. See also Kitsos and Toulias [12], and Kitsos *et al.* [13] for further reading. We recall its definition:

Definition 1. The p–dimensional random variable X follows the γ–order Normal distribution $\mathcal{N}^p_\gamma(\mu, \Sigma)$ with location parameter vector $\mu \in \mathbb{R}^{p\times 1}$ and positive definite scale matrix $\Sigma \in \mathbb{R}^{p\times p}$, when the density function, f_X, is of the form

$$f_X(x;\, \mu, \Sigma, \gamma) := C^p_\gamma |\det \Sigma|^{-1/2} \exp\left\{-\tfrac{\gamma-1}{\gamma} Q_\theta(x)^{\frac{\gamma}{2(\gamma-1)}}\right\}, \quad x \in \mathbb{R}^{p\times 1}, \tag{11}$$

with the p–quadratic form $Q_\theta(x) := (x-\mu)^{\mathrm{T}}\Sigma^{-1}(x-\mu)$, $x \in \mathbb{R}^{p\times 1}$, and $\theta := (\mu, \Sigma) \in \mathbb{R}^{(p\times 1)\times(p\times p)}$. We shall write $X \sim \mathcal{N}^p_\gamma(\mu, \Sigma)$. The normality factor C^p_γ is defined as

$$C^p_\gamma := \pi^{-p/2} \frac{\Gamma\left(\frac{p}{2}+1\right)}{\Gamma(p\frac{\gamma-1}{\gamma}+1)} \left(\tfrac{\gamma-1}{\gamma}\right)^{p\frac{\gamma-1}{\gamma}}. \tag{12}$$

Notice that, for $\gamma = 2$, $\mathcal{N}_2^p(\mu, \Sigma)$ is the well known multivariate normal distribution. It can be also easily noticed that the parameter vector μ is in fact the mean vector of the $\mathcal{N}_\gamma^p$ distribution, i.e. $\mu = \mathrm{E}[X]$ for all parameters $\gamma \in \mathbb{R} \setminus [0, 1]$.

Denote now with $\mathbb{E}_\theta$ the area of the p–ellipsoid $Q_\theta(x) \leq 1$, $x \in \mathbb{R}^{p\times 1}$. The family of $\mathcal{N}_\gamma^p(\mu, \Sigma)$, i.e. the family of the elliptically contoured γ–order Normals, provides a smooth bridging between the multivariate (and elliptically countered) Uniform, Normal and Laplace r.v. U, Z and L, i.e. between $U \sim \mathcal{U}^p(\mu, \mathit{\Sigma})$, $Z \sim \mathcal{N}^p(\mu, \Sigma)$ and Laplace $L \sim \mathcal{L}^p(\mu, \Sigma)$ r.v. as well as the multivariate degenerate Dirac distributed r.v. $D \sim \mathcal{D}^p(\mu)$ (with pole at the point μ). That is, the $\mathcal{N}_\gamma^p$ family of distributions, not only generalizes the usual Normal but also two other very significant distributions, as the Uniform and Laplace distributions, are induced. The above discussion is summarized in the following Theorem, see Kitsos *et al.* [13].

Theorem 1. *The elliptically contoured p–variate γ–order Normal distribution $\mathcal{N}_\gamma^p(\mu, \Sigma)$ for order values of $\gamma = 0, 1, 2, \pm\infty$ coincides with*

$$\mathcal{N}_\gamma^p(\mu, \Sigma) = \begin{cases} \mathcal{D}^p(\mu), & \textit{for} \;\; \gamma = 0 \;\; \textit{and} \;\; p = 1, 2, \\ 0, & \textit{for} \;\; \gamma = 0 \;\; \textit{and} \;\; p \geq 3, \\ \mathcal{U}^p(\mu, \Sigma), & \textit{for} \;\; \gamma = 1, \\ \mathcal{N}^p(\mu, \Sigma), & \textit{for} \;\; \gamma = 2, \\ \mathcal{L}^p(\mu, \Sigma), & \textit{for} \;\; \gamma = \pm\infty. \end{cases} \tag{13}$$

Considering the above Theorem, the definition values of the shape parameter γ of $\mathcal{N}_\gamma^p$ distributions can be extended to include the limiting extra values of $\gamma = 0, 1, \pm\infty$ respectively, i.e. γ can now be considered as a real number outside the open interval $(0, 1)$. Particularly, when $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \Sigma)$,

$$\mathcal{N}_0^p := \lim_{\gamma \to 0^-} \mathcal{N}_\gamma^p, \quad \mathcal{N}_1^p := \lim_{\gamma \to 1^+} \mathcal{N}_\gamma^p, \quad \mathcal{N}_{\pm\infty}^p := \lim_{\gamma \to \pm\infty} \mathcal{N}_\gamma^p. \tag{14}$$

Hence, the Uniform, Normal, Laplace and also the degenerate distribution $\mathcal{N}_0^p$ (like the Dirac for dimensions $p = 1, 2$), coinciding with N_1^p, N_2^p, $N_{\pm\infty}^p$ and N_0^p respectively, can be considered as members of the "extended" $\mathcal{N}_\gamma^p$ family of generalized Normal distributions, with $\gamma \in (\mathbb{R} \cup \{\pm\infty\}) \setminus (0,\, 1)$.

Notice also that $\mathcal{N}_1^1(\mu, \sigma)$ coincides with the known (continuous) Uniform distribution $\mathcal{U}(\mu - \sigma, \mu + \sigma)$. Specifically, for every Uniform distribution expressed with the usual notation $\mathcal{U}(a, b)$, it holds that $\mathcal{U}(a, b) = \mathcal{N}_1^1(\frac{a+b}{2}, \frac{b-a}{2}) = \mathcal{U}^1(\mu, \sigma)$. Also $\mathcal{N}_2(\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2)$, $\mathcal{N}_{\pm\infty}(\mu, \sigma^2) = \mathcal{L}(\mu, \sigma)$ and also $\mathcal{N}_0(\mu, \sigma) = \mathcal{D}(\mu)$. Therefore the following holds.

Corollary 1. *The univariate γ–ordered Normal distributions $\mathcal{N}_\gamma^1(\mu, \sigma^2)$ for order values $\gamma = 0, 1, 2, \pm\infty$ coincides with the usual (univariate) Dirac $\mathcal{D}(\mu)$, Uniform $\mathcal{U}(\mu - \sigma, \mu + \sigma)$, Normal $\mathcal{N}(\mu, \sigma^2)$ and Laplace $\mathcal{L}(\mu, \sigma)$ distributions respectively.*

For the multivariate normally distributed $X \sim \mathcal{N}^p(\mu, \Sigma)$ it is clear, from (11), that the maximum density value $\max f_X = f_X(\mu) = (2\pi)^{-p/2} |\det \Sigma|^{-1/2}$

decreases as dimension $p \in \mathbb{N}$ rises, providing "flattened" probability densities. This is also true for the multivariate Laplace distributed $X \sim \mathcal{L}^p(\mu, \Sigma) = \mathcal{N}^p_{\pm\infty}(\mu, \Sigma)$. In fact, for a Laplace distributed r.v. X it holds that $\max f_X = \pi^{-p/2} \frac{1}{p!} \Gamma(\frac{p}{2}+1) |\det \Sigma|^{-1/2}$ and therefore, the high–dimensional Laplace distributions densities are "flattened", since the maximum density values decreases as $p \in \mathbb{N}$ increases. This is true because, for dimensions $2p$, the maximum density is

$$\max f_X = C^p_{\pm\infty} |\det \Sigma|^{-1/2} = \pi^{-p/2} \frac{1}{(p+1)(p+2)...2p} |\det \Sigma|^{-1/2}.$$

Hence, as in the case of the Normal distribution, X obtains heavy tails as the dimension increases. However, this is not the case for the multivariate (and elliptically contoured) Uniform distributed $X \sim \mathcal{U}^p(\mu, \Sigma) = \mathcal{N}^p_1(\mu, \Sigma)$, because the volume of the corresponding p–elliptical–cylinder shape of their density functions, must always equal 1, although $\mathcal{U}^p$ have no tails to "absorb" probability mass when dimension increases, as the Normal or the Laplace distributions does. Based on the above discussion, the following Proposition shows that, among all elliptical multivariate Uniform distributions $\mathcal{U}^p(\mu, \Sigma)$ with fixed scale matrix Σ, the $\mathcal{U}^5(\mu, \Sigma)$ has the minimum $\max f_X$, see Kitsos *et al.* [13].

Proposition 1. *For the elliptically contoured Uniformly distributed $X \sim \mathcal{U}^p(\mu, \Sigma)$, we have*

$$\min_{p \in \mathbb{N}} \{\max f_X\} = \tfrac{15}{6\pi^2} |\det \Sigma|^{-1} = \max \mathcal{U}^5(\mu, \Sigma),$$

i.e. the 5–dimensional Uniform distribution provides the least of all maximum density values among all $\mathcal{U}^p(\mu, \Sigma)$ with fixed scale matrix Σ.

Recall now the cumulative distribution function (c.d.f.) $\Phi_Z(z)$ of the standardized normally distributed $Z \sim \mathcal{N}(0, 1)$, i.e.

$$\Phi_Z(z) = \tfrac{1}{2} + \tfrac{1}{2} \operatorname{erf}(\tfrac{z}{2}), \quad z \in \mathbb{R}, \tag{15}$$

with $\operatorname{erf}(\cdot)$ being the usual error function. Similarly, for the generalized $\mathcal{N}_\gamma$ family of distributions, the generalized error function $\operatorname{Erf}_{\gamma/(\gamma-1)}$, Gradshteyn and Ryzhik [6], is involved. Indeed, the following holds, see [8] for details.

Theorem 2. *Let X be a univariate γ–order normally distributed random variable, i.e. $X \sim \mathcal{N}^p_\gamma(\mu, \sigma^2)$ with p.d.f. f_X. Then the c.d.f. F_X of X is the c.d.f. Φ_Z of the standardized r.v. $Z = \frac{1}{\sigma}(X - \mu) \sim \mathcal{N}_\gamma(0, 1)$, and is given by*

$$\begin{aligned} F_X(x) = \Phi_Z(\tfrac{x-\mu}{\sigma}) &= \tfrac{1}{2} + \frac{\sqrt{\pi}}{2\Gamma(\frac{\gamma-1}{\gamma})\Gamma(\frac{\gamma}{\gamma-1})} \operatorname{Erf}_{\frac{\gamma}{\gamma-1}} \left\{ (\tfrac{\gamma-1}{\gamma})^{\frac{\gamma-1}{\gamma}} \left(\frac{x-\mu}{\sigma} \right) \right\} \\ &= 1 - \frac{1}{2\Gamma(\frac{\gamma-1}{\gamma})} \Gamma\left(\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma} \left(\frac{x-\mu}{\sigma} \right)^{\frac{\gamma}{\gamma-1}} \right), \quad x \in \mathbb{R}, \end{aligned} \tag{16}$$

with $\Gamma(\cdot, \cdot)$ being the upper (complementary) incomplete gamma function.

Applying the Shannon entropy on a γ–order normally distributed random variable the following Proposition holds.

Proposition 2. *The Shannon entropy of a random variable $X \sim \mathcal{N}_\gamma^p(\mu, \Sigma)$, with p.d.f. f_X, is of the form*

$$\mathrm{H}(X) = p\tfrac{\gamma-1}{\gamma} - \log\{C_\gamma^p |\det\Sigma|^{-1/2}\} = p\tfrac{\gamma-1}{\gamma} - \log\max f_X. \tag{17}$$

Proof. Let $C(\Sigma) = C(\Sigma; \gamma, p) := C_\gamma^p \det\Sigma|^{-1/2}$. From (11) and the definition (5) we have that the Shannon entropy of X is

$$\mathrm{H}(X) = -\log C(\Sigma) + C(\Sigma)\tfrac{\gamma-1}{\gamma}\int\limits_{\mathbb{R}^p} Q_\theta(x)^{\frac{\gamma}{2(\gamma-1)}} \exp\left\{-\tfrac{\gamma-1}{\gamma} Q_\theta(x)^{\frac{\gamma}{2(\gamma-1)}}\right\} dx.$$

Applying the linear transformation $z = (x-\mu)^{\mathrm{T}}\Sigma^{-1/2}$ with $dx = d(x-\mu) = \sqrt{|\det\Sigma|}dz$, the $\mathrm{H}(X)$ above is reduced to

$$\mathrm{H}(X) = -\log C(\Sigma) + C_\gamma^p \tfrac{\gamma-1}{\gamma}\int\limits_{\mathbb{R}^p} \|z\|^{\frac{\gamma}{\gamma-1}} \exp\left\{-\tfrac{\gamma-1}{\gamma}\|z\|^{\frac{\gamma}{\gamma-1}}\right\} dz.$$

Switching to hyperspherical coordinates, we get

$$\mathrm{H}(X) = -\log C(\Sigma) + C_\gamma^p \tfrac{\gamma-1}{\gamma}\omega_{p-1}\int\limits_{\mathbb{R}_+} \rho^{\frac{\gamma}{\gamma-1}} \exp\left\{-\tfrac{\gamma-1}{\gamma}\rho^{\frac{\gamma}{\gamma-1}}\right\} \rho^{p-1} d\rho,$$

where $\omega_{p-1} := 2\pi^{p/2}/\Gamma\left(\frac{p}{2}\right)$ is the volume of the $(p-1)$–sphere. Applying the variable change $du := d(\frac{\gamma-1}{\gamma}\rho^{\gamma/(\gamma-1)}) = \rho^{1/(\gamma-1)}d\rho$ we obtain successively

$$\begin{aligned}
\mathrm{H}(X) &= -\log C(\Sigma) + C_\gamma^p \omega_{p-1}\int\limits_{\mathbb{R}_+} ue^{-u}\rho^{\frac{(p-1)(\gamma-1)-1}{\gamma-1}} du \\
&= \log C(\Sigma) - C_\gamma^p \omega_{p-1}\int\limits_{\mathbb{R}_+} ue^{-u}\left(\rho^{\frac{\gamma}{\gamma-1}}\right)^{\frac{(p-1)(\gamma-1)-1}{\gamma}} du \\
&= -\log C(\Sigma) + C_\gamma^p \omega_{p-1}(\tfrac{\gamma}{\gamma-1})^{p\frac{\gamma-1}{\gamma}-1}\int\limits_{\mathbb{R}_+} u^{p\frac{\gamma-1}{\gamma}} e^{-u} du \\
&= -\log C(\Sigma) + p\tfrac{\gamma-1}{\gamma}\Gamma(p\tfrac{\gamma-1}{\gamma}) C_\gamma^p \omega_{p-1}.
\end{aligned}$$

Finally, by substitution of the volume ω_{p-1} and the normalizing factor $C(\Sigma)$ and C_γ^p, as in (12), relation (17) is obtained.

Corollary 2. *From Theorem 1 the Shannon entropy for the multivariate (and elliptically contoured) Uniform, Normal and Laplace distributions, i.e.*

$$\mathrm{H}(X) = \begin{cases} \log\frac{\pi^{p/2}}{\Gamma(\frac{p}{2}+1)}\sqrt{|\det\Sigma|}, & \textit{for } X \sim \mathcal{N}_1^p(\mu,\Sigma) = \mathcal{U}^p(\mu,\Sigma), \\ \frac{1}{2}\log\{(2\pi e)^p\,|\det\Sigma|\}, & \textit{for } X \sim \mathcal{N}_2^p(\mu,\Sigma) = \mathcal{N}^p(\mu,\Sigma), \\ p + \log\frac{p!\pi^{p/2}}{\Gamma(\frac{p}{2}+1)}\sqrt{|\det\Sigma|}, & \textit{for } X \sim \mathcal{N}_{\pm\infty}^p(\mu,\Sigma) = \mathcal{L}^p(\mu,\Sigma), \end{cases} \tag{18}$$

while for the univariate case $p = 1$, the Shannon entropy is reduced to

$$\mathrm{H}(X) = \begin{cases} \log 2\sigma, & \textit{for } X \sim \mathcal{N}_1^1(\mu, \sigma^2) = \mathcal{U}^1(\mu, \sigma^2) = \mathcal{U}(\mu - \sigma, \mu - \sigma), \\ \log \sqrt{2\pi e}\sigma, & \textit{for } X \sim \mathcal{N}_2^1(\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2), \\ 1 + \log 2\sigma, & \textit{for } X \sim \mathcal{N}_{\pm\infty}^1(\mu, \sigma^2) = \mathcal{L}^1(\mu, \sigma^2) = \mathcal{L}(\mu, \sigma). \end{cases}$$

where $\mathcal{U}(\mu - \sigma, \mu - \sigma)$, $\mathcal{N}(\mu, \sigma^2)$ and $\mathcal{L}(\mu, \sigma)$ are the usual notations for the univariate Uniform, Normal and Laplace distributions respectively.

Proof. Let $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \Sigma)$ and recall (14). Applying Theorem 1 into (17) the top branch of (18) for $\gamma = 1$ is obtained (in limit), i.e. $\mathrm{H}(X_1) := \lim_{\gamma \to 1^+} \mathrm{H}(X_\gamma)$, the middle branch of (18) for $\gamma = 2$ (normality), while the last branch of (18) is obtained for $\gamma = \pm\infty$ (in limit), i.e. $\mathrm{H}(X_{\pm\infty}) := \lim_{\gamma \to \pm\infty} \mathrm{H}(X_\gamma)$.

For the generalized α–Shannon entropy the following holds.

Proposition 3. *The α–Shannon entropy H_α of the multivariate $X \sim \mathcal{N}_\gamma(\mu, \Sigma)$ is given by*

$$\mathrm{H}_\alpha(X) = \tfrac{2\gamma - \alpha}{2\gamma} p + \tfrac{p}{2} \log \left\{ 2\pi (\tfrac{\alpha-1}{\alpha})^{\alpha-1} (\tfrac{\gamma}{\gamma-1})^{\alpha\frac{\gamma-1}{\gamma}} \left[\frac{\Gamma(p\frac{\gamma-1}{\gamma} + 1)}{\Gamma(p\frac{\alpha-1}{\alpha} + 1)} \right]^{\frac{\alpha}{p}} |\det \Sigma|^{\frac{\alpha}{2p}} \right\}. \tag{19}$$

For $\alpha = \gamma$ it is $\mathrm{H}_\gamma(X) = \frac{1}{2} \log\{(2\pi e)^p |\det \Sigma|^{\gamma/2}\}$. Moreover, for a random variable X following the multivariate Uniform, Normal and Laplace distributions ($\gamma = 1, 2, \pm\infty$ respectively), it is

$$\mathrm{H}_\alpha(X) = \begin{cases} \frac{2-\alpha}{2} p + h_{\gamma,a}^p, & X \sim \mathcal{N}_1^p(\mu, \Sigma), \\ p + \frac{\alpha}{2} \log\left\{(2/e)^{p/2} \Gamma(\frac{p}{2} + 1)\right\} + h_{\gamma,\alpha}^p, & X \sim \mathcal{N}_2^p(\mu, \Sigma), \\ p + \frac{p}{2} \log p! + h_{\gamma,a}^p, & X \sim \mathcal{N}_{\pm\infty}^p(\mu, \Sigma), \end{cases} \tag{20}$$

where $h_{\gamma,\alpha}^p := \frac{\alpha}{2} \log\{(2\pi)^{p/\alpha} (\frac{\alpha-1}{\alpha})^{p(\alpha-1)/\alpha} [\Gamma(p\frac{\alpha-1}{\alpha} + 1)]^{-1} \sqrt{|\det \Sigma|}\}$. For the limiting degenerate case of $\gamma = 0$ we obtain $\mathrm{H}_\alpha(X_0) = (\operatorname{sgn} \alpha)(+\infty)$, for $\alpha \neq 0$ while $\mathrm{H}_0(X_0) = p \log \sqrt{2\pi e}$.

Proof. Applying (17) to (10) and with some algebra, (19) holds. Recall Theorem 1. For the order values $\gamma = 1$, $\gamma = 2$ and $\gamma = \pm\infty$, the α–Shannon entropy H_α of the uniformly, normally and Laplace distributed $X_1 \sim \mathcal{U}^p(\mu, \Sigma)$, $X_2 \sim \mathcal{N}^p(\mu, \Sigma)$ and $X_{\pm\infty} \sim \mathcal{L}^p(\mu, \Sigma)$ respectively are given by (20). For the limiting case of $\gamma = 0$, we apply the Stirling's asymptotic formula, $k! \approx \sqrt{2\pi k} (\frac{k}{e})^k$ as $k \to \infty$, and we finally derive $\mathrm{H}_\alpha(X_0) = (\operatorname{sgn} \alpha)(+\infty)$.

We discuss now the Rényi entropy, which generalizes the Shannon entropy, and can be best introduced through the concept of *generalized random variables*. These variables extend the usual notion of a random experiment that cannot always be observed, see for details [16]. For a p–variate continuous random variable with p.d.f. f_X, the Rényi entropy $\mathrm{R}_\alpha(X)$ is defined, through the α–norm of $f_X \in \mathscr{L}^\alpha(\mathbb{R}^p)$, by

$$\mathrm{R}_\alpha(X) := -\tfrac{\alpha}{\alpha-1}\log\|f_X\|_\alpha = \tfrac{1}{1-\alpha}\log\int\limits_{\mathbb{R}^p}|f_X(x)|^\alpha dx, \quad \alpha\in\mathbb{R}_+^*\setminus 1, \tag{21}$$

where $\mathbb{R}_+^* := \{\alpha\in\mathbb{R} : \ \alpha > 0\}$. For the limiting case of $\alpha\to 1$ the Rényi entropy converges to the usual Shannon entropy $\mathrm{H}(X)$ as in (5). Notice that the minus sign is used in (21) to be in line with the definition of (5).

Considering a spherically contoured r.v. from the $\mathcal{N}_\gamma$ family of distributions, its Rényi entropy was given in [18]. We are extend now the corresponding results to the general multivariate, and hence elliptically contoured, case.

Theorem 3. *For the elliptically contoured γ–order normally distributed r.v. $X_\gamma\sim\mathcal{N}_\gamma^p(\mu,\Sigma)$, with p.d.f. f_{X_γ}, the Rényi R_α entropy of X_γ is given by*

$$\mathrm{R}_\alpha(X_\gamma) = p\tfrac{\gamma-1}{\gamma(\alpha-1)}\log\alpha - \log C_\gamma^p(\Sigma) = p\tfrac{\gamma-1}{\gamma(\alpha-1)}\log\alpha - \log\max f_{X_\gamma}, \tag{22}$$

for all the defined parameters $\alpha\in\mathbb{R}_+^*\setminus\{1\}$ and $\gamma\in\mathbb{R}\setminus[0,\,1]$.

Proof. Consider the p.d.f. f_{X_γ} as in (11). From the definition (21) it is

$$\mathrm{R}_\alpha(X_\gamma) = \tfrac{\alpha}{1-\alpha}\log C_\gamma^p(\Sigma) + \tfrac{1}{1-\alpha}\log\int\limits_{\mathbb{R}^p}\exp\left\{-\tfrac{\alpha(\gamma-1)}{\gamma}\left[(x-\mu)\Sigma^{-1}(x-\mu)^{\mathrm{T}}\right]^{\frac{\gamma}{2(\gamma-1)}}\right\}dx.$$

Applying the linear transformation $z = (x-\mu)\Sigma^{-1/2}$ with $dx = d(x-\mu) = \sqrt{|\det\Sigma|}dz$, the R_α above is reduced to

$$\mathrm{R}_\alpha(X_\gamma) = \tfrac{\alpha}{1-\alpha}\log C_\gamma^p(\Sigma) + \tfrac{1}{1-\alpha}\log\int\limits_{\mathbb{R}^p}\exp\left\{-\tfrac{\alpha(\gamma-1)}{\gamma}\|z\|^{\frac{\gamma}{\gamma-1}}\right\}dz.$$

Switching to hyperspherical coordinates, we get

$$\mathrm{R}_\alpha(X_\gamma) = \tfrac{\alpha}{1-\alpha}\log\left\{C_\gamma^p(\Sigma)\omega_{p-1}^{1/\alpha}\right\} + \tfrac{1}{1-\alpha}\log\int\limits_{\mathbb{R}_+}\exp\left\{-\tfrac{\alpha(\gamma-1)}{\gamma}\rho^{\frac{\gamma}{\gamma-1}}\right\}\rho^{p-1}d\rho,$$

where $\omega_{p-1} = 2\pi^{p/2}/\Gamma\left(\frac{p}{2}\right)$ is the volume of the $(p-1)$–sphere. Assuming $du := d(\frac{\gamma-1}{\gamma}\rho^{\gamma/(\gamma-1)}) = \rho^{1/(\gamma-1)}d\rho$ we obtain successively

$$\begin{aligned}
\mathrm{R}_\alpha(X_\gamma) &= \tfrac{\alpha}{1-\alpha}\log M(\Sigma) + \tfrac{1}{1-\alpha}\log\int\limits_{\mathbb{R}_+}e^{-\alpha u}\rho^{\frac{(p-1)(\gamma-1)-1}{\gamma-1}}du\\
&= \tfrac{\alpha}{1-\alpha}\log M(\Sigma) + \tfrac{1}{1-\alpha}\log(\tfrac{\gamma}{\gamma-1})^{p\frac{\gamma-1}{\gamma}-1} + \tfrac{1}{1-\alpha}\log\int\limits_{\mathbb{R}_+}e^{-\alpha u}u^{p\frac{\gamma-1}{\gamma}-1}du\\
&= \tfrac{\alpha}{1-\alpha}\log M(\Sigma) + \tfrac{1}{1-\alpha}\log(\tfrac{\gamma}{\gamma-1})^{p\frac{\gamma-1}{\gamma}-1} - p\tfrac{\gamma-1}{\gamma}\cdot\tfrac{\log\alpha}{1-\alpha} + \tfrac{1}{1-\alpha}\log\Gamma(p\tfrac{\gamma-1}{\gamma}),
\end{aligned}$$

where $M(\Sigma) := C^p_\gamma(\Sigma)\omega_{p-1}^{1/\alpha}$. Finally, by substitution of the volume ω_{p-1} we obtain, through the normalizing factor $C^p_\gamma(\Sigma)$ as in (12),

$$\mathrm{R}_\alpha(X_\gamma) = -\tfrac{\alpha}{\alpha-1}\log C^p_\gamma(\Sigma) + \tfrac{1}{\alpha-1}\log C^p_\gamma(\Sigma) + p\tfrac{\gamma-1}{\gamma}\cdot\tfrac{\log\alpha}{\alpha-1},$$

and thus (22) holds true. $\square$

For the limiting parameter values $\alpha = 0, 1, +\infty$ we obtain a number of results for other well known measures of entropy, applicable to Cryptography, as the Hartley entropy, the Shannon entropy, and min–entropy respectively, while for $\alpha = 2$ the collision entropy is obtained. Therefore, from Theorem 3, we have the following.

Corollary 3. *For the special cases of $\alpha = 0, 1, 2, +\infty$ the Rényi entropy of the elliptically contoured r.v. $X_\gamma \sim \mathcal{N}_\gamma(\mu, \Sigma)$ is reduced to*

$$\mathrm{R}_\alpha(X_\gamma) = \begin{cases} +\infty, & \textit{for } \alpha = 0, \quad (\textit{Hartley entropy}) \\ p\frac{\gamma-1}{\gamma} - \log\max f_{X_\gamma}, & \textit{for } \alpha = 1, \quad (\textit{Shannon entropy}) \\ p\frac{\gamma-1}{\gamma}\log 2 - \log\max f_{X_\gamma}, & \textit{for } \alpha = 2, \quad (\textit{collision entropy}) \\ -\log\max f_{X_\gamma}, & \textit{for } \alpha = +\infty, \quad (\textit{min–entropy}) \end{cases}$$

where $\max f_{X_\gamma} = C^p_\gamma(\Sigma)$.

Rényi entropy $\mathrm{R}_\alpha(X_\gamma)$, as in (22), is an decreasing function of parameter α and therefore

$$\mathrm{R}_{+\infty}(X_\gamma) < \mathrm{R}_2(X_\gamma) < \mathrm{R}_1(X_\gamma) < \mathrm{R}_0(X_\gamma), \quad \gamma \in \mathbb{R}\setminus[0\,1].$$

Example 1. For the multivariate (and elliptically contoured) Uniform random variable $U \sim \mathcal{U}(\mu, \Sigma)$, the Hartley, Shannon, collision and the min–entropy are coincide as,

$$\mathrm{R}_\alpha(U) = \log\frac{\pi^{p/2}\sqrt{|\det\Sigma|}}{\Gamma(\frac{p}{2}+1)}, \quad \alpha \in \mathbb{R}_+ \setminus 1,$$

while for the univariate case of $U \sim \mathcal{U}(a, b) = U(\mu-\sigma, \mu-\sigma)$ the Rényi entropy is reduced to $\mathrm{R}_\alpha(U) = \log(b-a) = \log(2\sigma)$, $\alpha \in \mathbb{R}_+ \setminus 1$.

Notice that for a uniformly distributed r.v. the Rényi entropy R_α is α–invariant, depending only on the dimension $p \in \mathbb{N}$ and the scale parameter σ.

Example 2. For the multivariate (and elliptically contoured) Laplace random variable $L \sim \mathcal{L}(\mu, \Sigma)$, the Hartley, Shannon, collision and the min– entropies are given by,

$$\mathrm{R}_\alpha(L) = \begin{cases} +\infty, & \alpha = 0, \quad \text{(Hartley entropy)} \\ p + \log\{p!\pi^{p/2}\sqrt{|\det\Sigma|}\Gamma(\frac{p}{2}+1)^{-1}\}, & \alpha = 1, \quad \text{(Shannon entropy)} \\ \log\{2^p p!\pi^{p/2}\sqrt{|\det\Sigma|}\Gamma(\frac{p}{2}+1)^{-1}\}, & \alpha = 2, \quad \text{(collision entropy)} \\ \log\{p!\pi^{p/2}\sqrt{|\det\Sigma|}\Gamma(\frac{p}{2}+1)^{-1}\}, & \alpha = +\infty. \quad \text{(min–entropy)} \end{cases}$$

The generalized Fisher's entropy type information of a random variable following the multivariate $\mathcal{N}_\gamma^p$, is given, see [11].

Theorem 4. *The generalized Fisher's information* J_α *of a r.v.* $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \lambda\Sigma^*)$ *where* $\lambda \in \mathbb{R}_+^*$ *and* Σ^* *is a real matrix with unit orthogonal vectors, i.e.* $\Sigma^* \in \mathbb{R}_\perp^{p\times p}$*, is given by*

$$\mathrm{J}_\alpha(X_\gamma) = \left(\tfrac{\gamma}{\gamma-1}\right)^{\frac{\alpha}{\gamma}} \frac{\Gamma\left(\frac{\alpha+p(\gamma-1)}{\gamma}\right)}{\lambda^{\alpha/2}\Gamma\left(p\frac{\gamma-1}{\gamma}\right)}, \quad \alpha \in \mathbb{R}_+ \setminus 1. \tag{23}$$

For the defined generalized Fisher's information measure and the γ–ordered Normal, it is clear that the values of $\mathrm{J}_\alpha(X_\gamma)$ depends on the two parameters α and γ. Therefore, we shall investigate for which values of α and γ the $\mathrm{J}_\alpha(X_\gamma)$ is bounded.

In the following Proposition we provide some inequalities for the generalized Fisher's entropy type information measure J_α for the family of the γ–order Normal distributions with positive order γ, i.e. for $\mathrm{J}_\alpha(X_\gamma)$ where $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \sigma^2\mathbb{I}_p)$, considering parameters $\alpha > 1$ and $\gamma > 2$. See [11] for details.

Proposition 4. *The generalized Fisher's information measure* J_α *of a random variable* X_γ *following the multivariate and spherically contoured* γ*–order Normal distribution, i.e.* $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \sigma^2\mathbb{I}_p)$, $\alpha, \gamma \geq 2$*, satisfy the inequalities*

$$\mathrm{J}_\alpha(X_\gamma) \begin{cases} > p\sigma^{-\alpha}, & \text{for } \alpha > \gamma, \\ = p\sigma^{-\alpha}, & \text{for } \alpha = \gamma, \\ < p\sigma^{-\alpha}, & \text{for } \alpha < \gamma. \end{cases} \tag{24}$$

Corollary 4. *The generalized Fisher's information* J_α *of a spherically contoured r.v.* $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \sigma^2\mathbb{I}_p)$*, with* $\alpha/\gamma \in \mathbb{N}^*$*, is reduced to*

$$\mathrm{J}_\alpha(X_\gamma) = \sigma^{-\alpha}(\gamma-1)^{-\alpha\gamma} \prod_{k=1}^{\alpha/\gamma} \{\alpha - p + (p-k)\gamma\}. \tag{25}$$

The following Fig. 1 depicts the generalized Fisher's information J_α of the spherically contoured bivariate (left sub–figure) and trivariate (right sub–figure) γ–order normally distributed random variables $X \sim \mathcal{N}_\gamma^p(\mu, \mathbb{I}_p)$, $p = 2, 3$, along the parameter $\alpha > 1$, and for various shape parameters $\gamma = 1, 1.1, \ldots, 1.9, 2, 3, \ldots, 10$. The usual Normal distribution case of $\gamma = 2$ is also highlighted.

3 Other $\mathcal{N}_\gamma$–based extensions

We present here some distributions that are based on the γ–order Normal distribution.

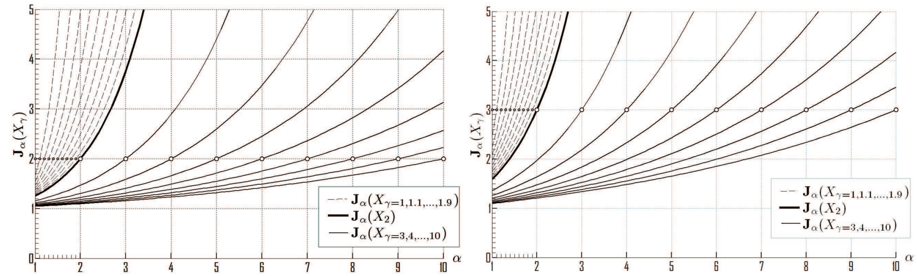


Fig. 1. Graphs of $\mathrm{J}_\alpha(X_\gamma)$ curves across parameter $\alpha > 1$, with $X_\gamma \sim \mathcal{N}_\gamma^p(\mu, \mathbb{I}_p)$, for $p = 2$ (left) and $p = 3$ (right), which are evaluated for various γ values.

3.1 Generalized Lognormal distribution

The Lognormal distribution is defined as the distribution of a random variable whose logarithm is normally distributed, and usually is formulated with two parameters. It is widely applied in life sciences, including Biology, Ecology, Geology, Meteorology as well as Economics, Finance, and Risk Analysis, see Crow and Shimizu [5]. Also, it plays an important role in Astrophysics and Cosmology, see Bernardeau and Kofman [2], Blasi *et al.* [3] among others.

Furthermore, Log–Uniform and Log–Laplace distribution can be similarly defined with applications in Finance, see Yan and Hanson [21], Kozubowski and Podgórski [14]. Especially, the power–tail phenomenon of the Log–Laplace distribution, Kozubowski and Podgórski [15], attracts, very often, attention in Environmental Sciences, Physics and Economics.

The Lognormal distribution can be also extended to the γ–order Lognormal distribution, denoted with $\mathcal{LN}(\mu, \sigma^2)$, in the sense that if $X \sim \mathcal{N}_\gamma^1(\mu, \sigma^2)$, with p.d.f. f_X, then $Y = e^X$ follows the $\mathcal{LN}_\gamma(\mu, \sigma)$ with p.d.f.

$$g_Y(y) := \tfrac{1}{y} f_X(\log y) = C_\gamma^1 (\sigma y)^{-1} \exp\left\{ -\tfrac{\gamma-1}{\gamma} \left| \frac{\log y - \mu}{\sigma} \right|^{\frac{\gamma}{\gamma-1}} \right\}, \tag{26}$$

where $y \in \mathbb{R}_+^* := \mathbb{R}_+ \setminus 0$. Moreover, if $X \sim \mathcal{LN}_\gamma(\mu, \sigma)$ then $\log X \sim \mathcal{N}_\gamma^1(\mu, \sigma^2)$.

For certain shape parameter values γ the $\mathcal{LN}_\gamma$ distribution is reduced to Log–Uniform and Log–Laplace distribution. See [19] for further reading on the subject.

3.2 Truncated γ–order Normal and Lognormal distribution

Let X be a univariate r.v. from $\mathcal{N}_\gamma(\mu, \sigma^2)$ with p.d.f. f_X, as in (11), and c.d.f. F_X, as in (16). Then, the r.v. X_ρ^+ is said to follow the *right–truncated γ–order Normal distribution at the point $\rho \in \mathbb{R}$*, when its p.d.f. $f_{X_\rho^+}$ is of the form

$$f_{X_\rho^+}(x) := \frac{f_X(x)}{F_X(\rho)} = \frac{(\frac{\gamma}{\gamma-1})^{1/\gamma} \exp\left\{ -\frac{\gamma-1}{\gamma} \left| \frac{x-\mu}{\sigma} \right|^{\frac{\gamma}{\gamma-1}} \right\}}{2\sigma\Gamma(\frac{\gamma-1}{\gamma}) - \sigma\Gamma\left(\frac{\gamma-1}{\gamma}, \frac{\gamma-1}{\gamma} \left(\frac{\rho-\mu}{\sigma} \right)^{\frac{\gamma}{\gamma-1}} \right)}, \quad x \leq \rho,$$

and zero when $x > \rho$, while the corresponding c.d.f. is then $F_{X_\rho^+}(x) = F_X(x)/F_X(\rho)$, $x \leq \rho$, and unity when $x > \rho$. It is clear that $\mathrm{Mode}(X_\rho^+) = \mu$. The mean of X_ρ^+ can then be evaluated as

$$\mu_{X_\rho^+} := \mathrm{E}[X_\rho^+] = \tfrac{1}{F_X(\rho)} \int\limits_{-\infty}^{\rho} x f_X(x) dx,$$

with f_X as in (11). Applying the transformation $z := \frac{x-\mu}{\sigma}$, we have

$$\mu_{X_\rho^+} = \mu + \tfrac{C_1^1 \sigma}{F_X(\rho)} \int\limits_{-\infty}^{\frac{\rho-\mu}{\sigma}} z \exp\left\{-\tfrac{\gamma-1}{\gamma}|z|^{\frac{\gamma}{\gamma-1}}\right\} dz,$$

through F_X as in (16). Applying now transformation $u = u(z) := \frac{\gamma-1}{\gamma}|z|^{\frac{\gamma}{\gamma-1}}$, and hence $du = (\operatorname{sgn} z)\frac{\gamma-1}{\gamma}|z|^{\frac{1}{\gamma-1}} dz$ with $\operatorname{sgn} z$ being the sign of z, we obtain

$$\mu_{X_\rho^+} = \mu - \frac{C_1^1 \sigma}{F_X(\rho)} (\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}} \Gamma\left(2\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma}\left|\frac{\rho-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right), \tag{27}$$

and substituting (16) into (27), we derive

$$\mu_{X_\rho^+} = \mu - \frac{\sigma(\frac{\gamma}{\gamma-1})^{3-\frac{2}{\gamma}} \Gamma\left(2\frac{\gamma-1}{\gamma}, \frac{\gamma-1}{\gamma}\left|\frac{\rho-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right)}{2\Gamma(\frac{\gamma-1}{\gamma}) - \Gamma\left(\frac{\gamma-1}{\gamma}, \frac{\gamma-1}{\gamma}\left(\frac{\rho-\mu}{\sigma}\right)^{\frac{\gamma}{\gamma-1}}\right)}. \tag{28}$$

Respectively, the r.v. X_τ^- is said to follow the *left–truncated γ–order Normal distribution at the point $\tau \in \mathbb{R}$*, when its p.d.f. $f_{X_\tau^-}$ is given by

$$f_{X_\tau^-}(x) := \frac{f_X(x)}{1 - F_X(\tau)} = -\frac{(\frac{\gamma}{\gamma-1})^{1/\gamma} \exp\left\{-\frac{\gamma-1}{\gamma}\left|\frac{x-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right\}}{\sigma\Gamma\left(\frac{\gamma-1}{\gamma}, \frac{\gamma-1}{\gamma}\left(\frac{\tau-\mu}{\sigma}\right)^{\frac{\gamma}{\gamma-1}}\right)}, \quad x \geq \tau,$$

and zero for $x < \tau$, while the corresponding c.d.f. is then $F_{X_\tau^-}(x) = F_X(x)/[1-F_X(\tau)]$, $x \geq \tau$, and zero for $x < \tau$. Note also that $\mathrm{Mode}(X_\tau^-) = \mu$. Similarly to the right–truncated case, the mean of X_τ^- can then be evaluated as

$$\mu_{X_\tau^-} := \mathrm{E}[X_\tau^-] = \tfrac{1}{1-F_X(\tau)} \int\limits_{\tau}^{+\infty} x f_X(x) dx.$$

Applying again the transformation $z := \frac{x-\mu}{\sigma}$, we have

$$\mu_{X_\tau^-} = \mu + \tfrac{C_1^1 \sigma}{1-F_X(\tau)} \int\limits_{\frac{\tau-\mu}{\sigma}}^{+\infty} z \exp\left\{-\tfrac{\gamma-1}{\gamma}|z|^{\frac{\gamma}{\gamma-1}}\right\} dz,$$

through (16). Applying now transformation $u := u(z)$ we obtain

$$\mu_{X_\tau^-} = \mu + \tfrac{C_1^1\sigma}{1-F_X(\tau)}(\tfrac{\gamma}{\gamma-1})^{2\frac{\gamma-1}{\gamma}}\Gamma\left(2\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma}\left|\frac{\tau-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right), \tag{29}$$

and substituting (16) into (29), we derive

$$\mu_{X_\tau^-} = \mu - \frac{\sigma(\tfrac{\gamma}{\gamma-1})^{3-\frac{2}{\gamma}}\Gamma\left(2\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma}\left|\dfrac{\tau-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right)}{\Gamma\left(\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma}\left(\dfrac{\tau-\mu}{\sigma}\right)^{\frac{\gamma}{\gamma-1}}\right)}. \tag{30}$$

For the generalized Lognormal r.v. $Y := e^X \sim \mathcal{LN}_\gamma(\mu, \sigma^2)$, through (26), the r.v. Y_ρ^+ is said follows the *right–truncated γ–order Lognormal distribution at the (log–scaled) point $\rho \in \mathbb{R}_+^*$*, when its p.d.f. $g_{Y_\rho^+}$ is of the form

$$g_{Y_\rho^+}(y) := \frac{f_X(\log y)}{yF_X(\log\rho)} = \frac{(\tfrac{\gamma}{\gamma-1})^{1/\gamma}\exp\left\{-\tfrac{\gamma-1}{\gamma}\left|\dfrac{\log y-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right\}}{2\sigma y\Gamma(\tfrac{\gamma-1}{\gamma}) - \sigma y\Gamma\left(\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma}\left(\dfrac{\log\rho-\mu}{\sigma}\right)^{\frac{\gamma}{\gamma-1}}\right)},$$

when $0 < y \leq \rho$ and zero when $y > \rho$, while the corresponding c.d.f. is then $G_{Y_\rho^+}(y) = F_X(\log y)/[F_X(\log\rho)]$, $0 < y \leq \rho$, and unity when $y > \rho$.

Respectively, the r.v. Y_τ^- is said to follow the *left–truncated γ–order Normal distribution at the (log–scaled) point $\tau \in \mathbb{R}_+^*$*, when its p.d.f. $g_{Y_\tau^-}$ is given by

$$g_{Y_\tau^-}(y) := \frac{f_X(\log y)}{y - yF_X(\log\tau)} = -\frac{(\tfrac{\gamma}{\gamma-1})^{1/\gamma}\exp\left\{-\tfrac{\gamma-1}{\gamma}\left|\dfrac{\log y-\mu}{\sigma}\right|^{\frac{\gamma}{\gamma-1}}\right\}}{\sigma\Gamma\left(\tfrac{\gamma-1}{\gamma}, \tfrac{\gamma-1}{\gamma}\left(\dfrac{\log\tau-\mu}{\sigma}\right)^{\frac{\gamma}{\gamma-1}}\right)}, \quad y \geq \tau,$$

and zero when $0 < y < \tau$, while the corresponding c.d.f. is then $G_{Y_\tau^-}(y) = F_X(\log y)/[1 - F_X(\log\tau)]$, $y \geq \tau$, and zero when $0 < y < \tau$.

3.3 Asymmetric form of the γ–order Normal Distribution

A two–way asymmetric form of the $\mathcal{N}_\gamma$ family can be constructed using a pair (σ_1, σ_2) of asymmetric scale parameters, and/or a pair (γ_1, γ_2) of asymmetric shape parameters.

A r.v. X follows the (two–way) *asymmetric (γ_1,γ_2)–order Normal distribution*, denoted by $\mathcal{AN}_{\gamma_1,\gamma_2}(\mu, \sigma_1^2, \sigma_2^2)$, when its p.d.f. f_X is defined as

$$f_X(x) := \begin{cases} C\exp\left\{-\tfrac{\gamma_1-1}{\gamma_1}\left|\dfrac{x-\mu}{\sigma_1}\right|^{\frac{\gamma_1}{\gamma_1-1}}\right\}, & x < \mu, \\ C\exp\left\{-\tfrac{\gamma_2-1}{\gamma_2}\left|\dfrac{x-\mu}{\sigma_2}\right|^{\frac{\gamma_2}{\gamma_2-1}}\right\}, & x \geq \mu, \end{cases}$$

with μ being the (maximum density) turning point, and C being the (mutual) normalizing factor

$$C := \frac{2C^1_{\gamma_1} C^1_{\gamma_2}}{\sigma_1 C^1_{\gamma_2} + \sigma_2 C^1_{\gamma_1}} = \frac{1}{\sigma_1 \Gamma(\frac{\gamma_1 - 1}{\gamma_1})(\frac{\gamma_1 - 1}{\gamma_1})^{1/\gamma_1} + \sigma_2 \Gamma(\frac{\gamma_2 - 1}{\gamma_2})(\frac{\gamma_2 - 1}{\gamma_2})^{1/\gamma_1}}.$$

Equivalently, when X_1, X_2 are two univariate (and symmetric) r.v. from $\mathcal{N}_{\gamma_1}(\mu, \sigma_1^2)$ and $\mathcal{N}_{\gamma_2}(\mu, \sigma_2^2)$ with p.d.f. f_{X_1} and f_{X_2}, the asymmetric r.v. $X \sim \mathcal{AN}_{\gamma_1,\gamma_2}(\mu, \sigma_1^2, \sigma_2^2)$, adopts a p.d.f. given by

$$f_X(x) = \frac{2}{1 + \left(\frac{\sigma_2 C^1_{\gamma_2}}{\sigma_1 C^1_{\gamma_1}} \right)^{\operatorname{sgn}(\mu - x)}} \times \begin{cases} f_{X_1}(x), & x < \mu, \\ f_{X_2}(x), & x \geq \mu. \end{cases}$$

where $\operatorname{sgn}(\cdot)$ is the sign function. See Fig. 2 for the visualization of some "shape–asymmetric" $\mathcal{AN}_{2;\gamma}(0, 1, 1)$ distributions (with fixed shape parameter $\gamma_1 = 2$ for the left part of the p.d.f. and various γ parameter values for the right part of the asymmetric p.d.f.).

The c.d.f. of the asymmetric r.v. X as above is then given by

$$F_X(x) = \begin{cases} \frac{1+\operatorname{sgn}(x-\mu)}{1+k} + \frac{\operatorname{sgn}(x-\mu)}{(1+k)\Gamma(\frac{\gamma-1}{\gamma})} \Gamma\left(\frac{\gamma-1}{\gamma}, \frac{\gamma-1}{\gamma} \left| \frac{x-\mu}{\sigma} \right|^{\frac{\gamma}{\gamma-1}} \right), & x < \mu, \\ \frac{1}{1+k} + \frac{1}{1+k^{-1}} \left[2 - \Gamma(\frac{\gamma-1}{\gamma})^{-1} \Gamma\left(\frac{\gamma-1}{\gamma}, \frac{\gamma-1}{\gamma} \left(\frac{x-\mu}{\sigma} \right)^{\frac{\gamma}{\gamma-1}} \right) \right], & x \geq \mu, \end{cases}$$

where $k := (\sigma_2 C^1_{\gamma_2})(\sigma_1 C^1_{\gamma_1})^{-1}$.

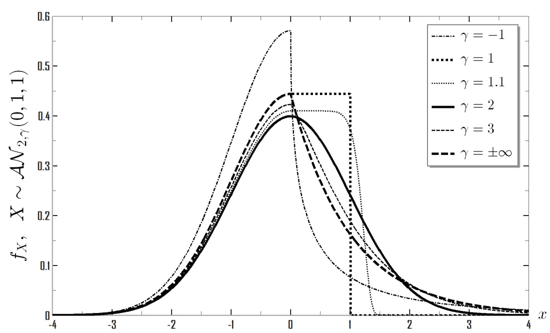


Fig. 2. Graph of the densities f_X, with $X \sim \mathcal{AN}_{2;\gamma}(0, 1, 1)$, for various "right" shape parameter values γ.

4 Discussion

In this paper we considered an exponential–power generalized form of the multivariate normal distribution, namely the γ–order Normal distribution, $\mathcal{N}^p_\gamma$.

This generalization was obtained through the study of the generalized entropy power, Kitsos and Tavoularis [9]. Two generalized Shannon entropies were evaluated and discussed (including the specific cases of the multivariate Uniform, Normal and Laplace distributions) for the multivariate (and elliptically distributed) γ–order normally distributed random variables. Moreover, the generalized entropy type information measure, J_α, which extends the known entropy type Fisher's information, was investigated through these random variables and certain boundaries of the J_α were obtained. Finally, three univariate $\mathcal{N}_\gamma$–based extensions were also given, i.e. the γ–order Lognormal distribution $\mathcal{LN}_\gamma$, the left/right truncated cases of $\mathcal{N}_\gamma$ and $\mathcal{LN}_\gamma$ distributions, and a two–way asymmetric $\mathcal{N}_\gamma$ distribution.

References

1. L.F. Bauer. *Kryptologie, Methoden und Maximen*, Springer–Verlag, London (1994)
2. F. Bernardeau and L. Kofman. Properties of the cosmological density distribution function, *Monthly Notices of the Royal Astrophys. J.*, 443, 479–498, 1995.
3. P. Blasi, S. Burles and A.V. Olinto. Cosmological magnetic field limits in an inhomogeneous Universe. *The Astrophysical Journal Letters*, 2, L79–L82, 1999.
4. T.M. Cover and J.A. Thomas. *Elements of Information Theory*, 2nd Ed., Wiley, 2006.
5. E.L. Crow and K. Shimizu. *Lognormal Distributions, Theory and Applications*, M. Dekker, New York & Basel, 1998.
6. I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products*, Elsevier, 2007.
7. J. Hoffstein, J. Pipher and J.H. Silverman. *An Introduction to Mathematical Cryptography*, Springer, 2008.
8. C.P. Kitsos, V.G. Vassiliadis and T.L. Toulias. MLE for the -order generalized normal distribution, *Discussiones Mathematicae Probability and Statistics*, 34, 1-2, 143–158, 2014.
9. C.P. Kitsos and N.K. Tavoularis. Logarithmic Sobolev inequalities for information measures, *IEEE Trans. Inform. Theory*, 55, 6, 2554–2561, 2009.
10. C.P. Kitsos and N.K. Tavoularis. New entropy type information measures, in *Information Technology Interfaces (ITI)*, V. Luzar–Stiffer, Z. Jarec and Z. Bekic (Eds), 255–259, 2009.
11. C.P. Kitsos and T.L. Toulias. Inequalities for the generalized entropy type Fisher's information, in *Handbook f Functional Equations: Functional Inequalities*, T.M. Rassias (Ed.), Springer, 2014.
12. C.P. Kitsos and T.L. Toulias. New information measures for the generalized normal distribution, *Information*, 1, 13–27, 2010.
13. C.P. Kitsos, T.L. Toulias and C.P. Trandafir. On the multivariate γ–ordered normal distribution, *Far East J. of Theoretical Statistics*, 38, 1, 49–73, 2012.
14. T.J. Kozubowski and K. Podgórski. Asymmetric Laplace laws and modeling financial data, *Math. Comput. Modelling*, 34, 1003–1021, 2000.
15. T.J. Kozubowski and K. Podgórski. Asymmetric Laplace distributions, *Math. Sci.*, 25, 37–46, 2001.
16. A. Rényi. *Probability Theory*, North–Holland (Ser. Appl. Math. Mech.), Amsterdam, 1970.
17. D.R. Stinson. *Cryptography: Theory and Practice*, 3rd Ed., CRC Press, 2006.

18. T.L. Toulias and C.P. Kitsos. Entropy measures and generalized Fisher's information, in *Stochastic Modelling and Data Analysis (SMTDA)*, e-Proceedings, 729–738, 2014.
19. T.L. Toulias and C.P. Kitsos. On the generalized Lognormal distribution, *J. Prob. and Stat.*, vol. 2013, 1–16, 2013.
20. I. Vajda. $\mathcal{X}^2$–divergence and generalized Fisher's information, in *Transactions of the 6th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, 873–886, 1973.
21. G. Yan and F.B. Hanson. Option pricing for a stochastic–volatility jump-diffusion model with log–uniform jump–amplitudes, in *Proceedings of the American Control Conference*, 2006. (DOI: 10.1109/ACC.2006.1657175)

S-weighted estimators

Jan Ámos Víšek[1]

Institute of Economic Studies,
Faculty of Social Siences, Charles University
Opletalova ulice 26, 110 01 Praha 1, the Czech Republic
(E-mail: visek@fsv.cuni.cz)

Abstract. An idea of implicit estimating the scale of disturbances, appearing in the *least median of squares*, the *least trimmed squares*, and the *least weighted squares*, found its explicit expression in the S-estimators. A similar idea of weighting order statistics of squared residuals, also implicitly employed in the *least median of squares* and the *least trimmed squares*, was in full generality used by the *least weighted squares*. A combination of these two idea yielded a new, flexible type of methods - the S-*weighted estimators*. Their properties, especially consistency is studied.
Keywords: Robust estimation of regression model, consistency of the S-weighted estimator under heteroscedasticity.

1 AN OVERVIEW OF PROBLEM

Many types of today already classical type of robust estimators, due to their dependence on the scale of disturbances, have lost affine equivariance. One way of healing it was brought by idea of implicit estimation of scale by the *least median of squares* (LMS) (Rousseeuw (1984)), the *the least trimmed squares*(LTS) (Hampel et al. (1986)) and by the *least weigted squares* (LWS) (Víšek (2000)). The idea found its explicit form in S-estimators (Rousseeuw, Yohai (1984)). Let's recall that moreover these estimators (together with Siegel's repeated median, Siegel (1982)) answered the challenge by Peter Bickel (1875) about possibilty of 50% breakdown point of an estimator of regression model. Moreover, S-estimators improved the efficiency of estimation. On the other hand, the high breakdown point estimators - especially those with zero-one weights - can (and typically do) suffer by a high sensitivity to an (arbitrarily) small shift of even one observation - see Figure 1 below. Then a smooth depressing the influential observations can be a remedy of this disadvantage. The later idea was worked up by the *least weighted squares*, Víšek (2000, 2011a). Combining the former and the later ideas, namely combining the S-*estimators* with the *least weighted squares* brings a flexible family of estimators which are studied in the rest of paper.

16^{th} *ASMDA Conference Proceedings, 30 June – 4 July 2015, Piraeus, Greece*
Research was supported by the Czech Science Foundation project No. 13-01930S Robust methods for nonstandard situations, their diagnostics and implementations.

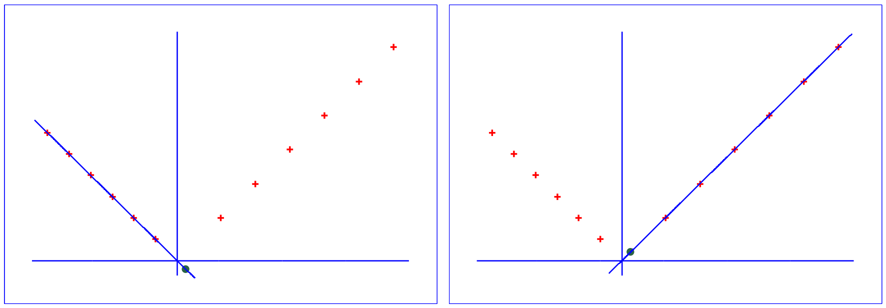


Figure 1
Notice that in both cases the estimator explains the most of data although the difference between the left and the right part of figure is a position of small bullet (close to origin) which is shifted from position below the axe (in the left part) x to the position above the axe x (in the right part).

2 RECALLING BASIC FRAMEWORK

Let $\mathcal{N}$ denote the set of all positive integers, R the real line and R^p the p-dimensional Euclidean space. All vectors will be assumed to be the column ones and throughout the paper we assume that all r.v.'s are defined on a basic probability space $(\Omega, \mathcal{A}, P)$, say. For a sequence of $(p+1)$-dimensional random variables (r. v.'s) $\{(X_i', e_i)'\}_{i=1}^{\infty}$, for any $n \in \mathcal{N}$ and a fixed $\beta^0 \in R^p$ the linear regression model given as

$$Y_i = X_i'\beta^0 + e_i = \sum_{j=1}^{p} X_{ij}\beta_j^0 + e_i, \quad i = 1, 2, ..., n \quad \text{or} \quad Y = X\beta^0 + e \qquad (1)$$

where $Y = (Y_1, Y_2, ..., Y_n)'$, $X = (X_1, X_2, ..., X_n)'$ and $e = (e_1, e_2, ..., e_n)'$, will be considered.

Prior to recalling some estimators let us give the assumptions which we adopt hereafter.

Conditions on explanatory variables and on disturbances
(C1) *The sequence* $\{(X_i', e_i)'\}_{i=1}^{\infty}$ *is sequence of independent* $(p+1)$*-dimensional random variables (r.v.'s) with distribution function* $F_{X,e_i}(v, u) = F^{(1)}(v^{(1)}) \cdot F^{(2)}_{X,e_i}(v, u)$ *where* $F^{(1)}(v^{(1)}) : R^1 \to [0, 1]$ *is d. f. degenerated at 1 and* $F^{(2)}_{X,e_i}(x, r) = F_X(x) \cdot F_{e_i}(r)$ *where* $F_{e_i}(r) = F_e(r\sigma_i^{-1})$ *with* $\mathbb{E}e_i = 0$, $\text{var}\,(e_i) = \sigma_i^2$ *and* $0 < \liminf_{i\to\infty} \sigma_i \leq \limsup_{i\to\infty} \sigma_i < \infty$. *Moreover,* $F_e(r)$ *(a parent d. f.) is absolutely continuous with density* $f_e(r)$ *bounded by* U_e. *Finally, there is* $q > 1$ *so that* $\mathbb{E}\,\|X_1\|^{2q} < \infty$ *(as* $F_X(x)$ *doesn't depend on i, the sequence* $\{X_i\}_{i=1}^{\infty}$ *is sequence of independent and identically distributed (i.i.d.) r.v.'s).*

Condition on objective function (see Rousseeuw, Yohai (1984), p.260.)

(R) $\rho : (0, \infty) \rightarrow (0, \infty)$, $\rho(0) = 0$, strictly increasing on $(0, c)$ and constant otherwise, symmetric and continuously differentiable (denote the derivative by ψ).

Remark 1. As $\rho(0) = 0$, due to symmetry of ρ and its differentiability, $\psi(0)$ is antisymmetric and $\psi(0) = 0$.

Let for any $\beta \in R^p$ define the i-th residual as $r_i(\beta) = Y_i - X_i^{'}\beta$ and $r^2_{(h)}(\beta)$ the h-th order statistic among the squared residuals, i.e. we have

$$r^2_{(1)}(\beta) \leq r^2_{(2)}(\beta) \leq \ \ldots \ \leq r^2_{(n)}(\beta). \tag{2}$$

Definition 1. Let (R) hold. Then

$$\hat{\beta}^{(S,\rho,n)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ \ : \ \sum_{i=1}^{n} \rho\left(\frac{r_i(\beta)}{\sigma}\right) = b \right\} \tag{3}$$

where $b = I\!E\rho(\frac{\varepsilon_1}{\sigma_0})$, is called the <u>*S-estimator*</u>, see Rousseeuw, Yohai (1984).

Remark 2. Putting $\tilde{\rho}(x) = \rho(\sqrt{x})$ we can write (3) as

$$\hat{\beta}^{(S,\tilde{\rho},n)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ \ : \ \sum_{i=1}^{n} \tilde{\rho}\left(\frac{r_i^2(\beta)}{\sigma^2}\right) = b \right\}, \tag{4}$$

we will need it a bit later.

In what follows we will consider a bit modified condition (R) including :
Conditions on weight and objective function
(WR)
- $w : [0, 1] \rightarrow [0, 1]$ is a continuous, non-increasing weight function with $w(0) = 1$. Moreover, w is Lipschitz in absolute value, i.e. there is L such that for any pair $u_1, u_2 \in [0, 1]$ we have $|w(u_1) - w(u_2)| \leq L \cdot |u_1 - u_2|$.
- $\rho : (0, \infty) \rightarrow (0, \infty)$, $\rho(0) = 0$, non-decreasing on $(0, \infty)$, symmetric and differentiable (denote the derivative by ψ).
- $\psi(r)/r$ is non-increasing for $r \geq 0$ with $\lim_{r \to 0_+} \frac{\psi(r)}{r} = 1$.

Notice that (WR) allow a bit wider family of ρ-functions than (R). Exactly as the *weighted least squares* are generalization of the *ordinary least squares* (OLS), we can have the *weighted S-estimators*, just plugging the weights between the summation sign and $\tilde{\rho}$ in (4). We obtain

$$\hat{\beta}^{(WS,\tilde{\rho},n)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ \ : \ \sum_{i=1}^{n} w_i \tilde{\rho}\left(\frac{r_i^2(\beta)}{\sigma^2}\right) = b \right\} \tag{5}$$

Finally, we can generalize the *weighted S-estimators* to the *S-weighted estimator* in the same way as the *weighted least squares* were generalized to the *least weighted squares*, the details are given below.

3 S-WEIGHTED ESTIMATOR AND ITS CONSISTENCY

Definition 2. Let (WR) hold. Then

$$\hat{\beta}^{(SW,n,w,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma(\beta) \in R^+ \; : \; \frac{1}{n} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \rho\left(\frac{r^2_{(i)}(\beta)}{\sigma^2}\right) = b \right\} \tag{6}$$

where $b = I\!\!E\rho(\frac{\varepsilon_1^2}{\sigma_0^2})$, is called the *S-weighted estimator*.

The name *S-weighted estimator* - by the order of words - should indicate, similarly as in the case of LMS, LTS or LWS, that the weights are prescribed to the order statistics of squared residuals rather than the squared residuals directly. This "trick" in fact allows to the method to decide itself which weight will be assigned to which observation.

Definition 3. Let $w : [0,1] \to [0,1]$ is a weight function. Then the solution of the extremal problem

$$\hat{\beta}^{(LWS,n,w)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r^2_{(i)}(\beta) \tag{7}$$

is called the *Least Weighted Squares* (LWS[1]), (Víšek (2000), (2011a) and references given there).

Remark 3. Let $\frac{n}{2} < h \le n$ and define two weight functions as follows:

$$w_{LMS}(r) = 1 \;\text{ for }\; r = \frac{h-1}{n}, \qquad w_{LMS}(r) = 0 \;\text{ otherwise}$$

and

$$w_{LTS}(r) = 1 \;\text{ for }\; r \le \frac{h-1}{n}, \qquad w_{LTS}(r) = 0 \;\text{ otherwise.}$$

Then we have for the weight functions $w_{LMS}(r)$ and $w_{LTS}(r)$

$$\hat{\beta}^{(LWS,n,w_{LMS})} = \hat{\beta}^{(LMS,n,h)} \qquad \textit{and} \qquad \hat{\beta}^{(LWS,n,w_{LTS})} = \hat{\beta}^{(LTS,n,h)},$$

respectively, see Rousseeuw (1984) and Hampel et al. (1986). Notice that the requirement (WR) allow for $\rho(r) = r^2$ so that $\hat{\beta}^{(LWS,n,w)}$ are special case of $\hat{\beta}^{(SW,n,w,\rho)}$.

Remark 4. When considering the optimality of selecting the weight function w, a very first idea - especially in the case when data contain (evidently several) leverage points - is to select w so that:

$$w(r) = \begin{cases} 1 & r \in [0, h/n] \\ 0 & r \in [g/n, 1] \end{cases}$$

[1]Notice please that the LWS assigns - similarly as the above defined S-weighted estimator - the weights to the order statistics of squared residuals, the classical *weighted least squares* (WLS) do it to the squared residuals directly and that is why WLS are not (generally) robust.

and decreasing linearly between h/n andg/n, see Figure 2 below. Numerical simulations showed that (exactly) optimal appeared a function given in the right part of Figure 2. On the other hand, the difference in efficiency (measured by means of MSE computed from the simulated data) between the left and right weight function was about 3% which indicate that the function w is rather flexible. However, what was (naturally) crucial, was the value of $1-g/n$ which has to be larger that k/n if we have k leverage points. Because we do not know the number of leverage points we have to start with a small value of g (may be a bit larger than $\frac{n}{2}$) and we may perform - due to the speed of computation of the estimates, see e. g. Víšek (2006b): - something like *Forward Search*, see Atkinson, Riani (2000).

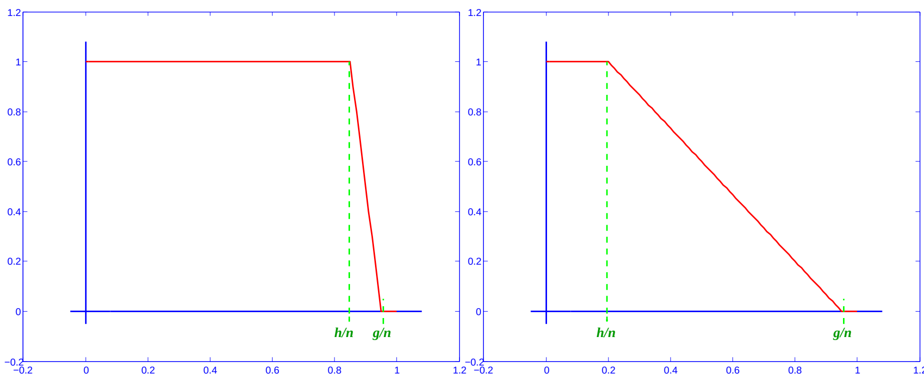


Figure 2

Remark 5. Denoting for any $\beta \in R^p$

$$\sigma = \sigma(\beta) = \sqrt{\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) r_{(i)}^2(\beta)},$$

we obtain from (7)

$$\hat{\beta}^{(LWS,n,w,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \frac{r_{(i)}^2(\beta)}{\sigma^2} = 1 \right\}. \quad (8)$$

Then putting $\rho(r) = r$, we can rewrite (8) as

$$\hat{\beta}^{(LWS,n,w,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \rho\left(\frac{r_{(i)}^2(\beta)}{\sigma^2}\right) = 1 \right\}.$$

Now, let us allow ρ to be an arbitrary function fulfilling the condition (R) and put $b = I\!E\left\{w(r)\rho(\frac{r^2}{\sigma^2})\right\}$. We arrive at (notice please an enlarged notation for the next estimator - enlarged about a one superindex "ρ", i. e. we write here $\hat{\beta}^{(LWS,n,w,\rho)}$ instead of what we have used in previous $\hat{\beta}^{(LWS,n,w)}$)

$$\hat{\beta}^{(LWS,n,w,\rho)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ : \sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) \rho\left(\frac{r_{(i)}^2(\beta)}{\sigma^2}\right) = b \right\}. \quad (9)$$

Evidently, (9) is the same as (6). So, we can interpret the *S-weighted estimators* also as a generalization of LWS. On the other hand, LWS is not a special case of the *S-estimators*, see the next remark.

Remark 6. We can rewrite (4) as

$$\hat{\beta}^{(S,\tilde{\rho},n)} = \arg\min_{\beta \in R^p} \left\{ \sigma \in R^+ \, : \, \sum_{i=1}^{n} \tilde{\rho} \left(\frac{r^2_{(i)}(\beta)}{\sigma^2} \right) = b \right\}. \tag{10}$$

Comparison of (8) and (10) implies that if we would want to prove that LWS can be represented as S-estimator, we have to show that for any weight function w there is a nondecrasing function ρ^* such that

$$\rho^* \left(\frac{r^2_{(i)}(\beta)}{\sigma^2} \right) = w \left(\frac{i-1}{n} \right) \frac{r^2_{(i)}(\beta)}{\sigma^2} \, . \tag{11}$$

Now, let us realize that $r^2_{(i)}(\beta) \le r^2_{(i+1)}(\beta)$ implies - due to the assumption that ρ^* is nondecrasing on $(0, \infty)$

$$\rho^* \left(\frac{r^2_{(i)}(\beta)}{\sigma^2} \right) \le \rho^* \left(\frac{r^2_{(i+1)}(\beta)}{\sigma^2} \right)$$

while we can have

$$w \left(\frac{i-1}{n} \right) \frac{r^2_{(i)}(\beta)}{\sigma^2} \le w \left(\frac{i}{n} \right) \frac{r^2_{(i+1)}(\beta)}{\sigma^2}$$

as well as

$$w \left(\frac{i-1}{n} \right) \frac{r^2_{(i)}(\beta)}{\sigma^2} \ge w \left(\frac{i}{n} \right) \frac{r^2_{(i+1)}(\beta)}{\sigma^2}.$$

It indicates that we cannot find the function ρ^* fulfilling the assumptions (R) and simultaneously (10). So the conclusion is: The LWS is not a special case of S-estimator but - taking into account once again just performed considerations - we can also conclude that S-estimators cannot be represented as LWS by a special adjustment of the weights w_i's.

Now, let's recall that the *least weighted squares* $\hat{\beta}^{(LWS,n,w)}$ are evidently the *weighted least squares* $\hat{\beta}^{(WLS,n,w)}$ applied on a permutation, say π, of r. v.'s from (1). The permutation in question of course depends on $\omega \in \Omega$, i. e. the permutation π is a mapping $\pi : \Omega \to \{1, 2, ..., n\}$, for technicalities see (Víšek (2011a)). In the same way we can show that the *S-weighted estimator* $\hat{\beta}^{(SW,n,w,\rho)}$ is also the *weighted S-estimator* $\hat{\beta}^{(WS,n,w,\rho)}$ for an appropriate permutation of data, the permutation which depends again on $\omega \in \Omega$. It immediately proves the existence of a solution of the extremal problem (6). Of course, at thi point we need the existence of a solution of the corresponding extremal problem for $\hat{\beta}^{(WS,n,w,\rho)}$ but it follows by the same arguments which Rousseeuw and Yohai (1984) applied in the proof of existence $\hat{\beta}^{(S,n,w,\rho)}$.

Now, we are going to repeat (in a bit modified form) the arguments of Peter Rousseeuw and Victor Yohai (1984). As $\hat{\sigma}^2$ is given (as by-product) by Definition 2 (as a minimum), its partial derivatives with respect to β are equal to zero. Moreover, for any $\beta \in R^p$

$$\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right)\rho\left(\frac{r^2_{(i)}(\beta)}{\sigma^2}\right) \geq b \tag{12}$$

because otherwise for such a β for which (12) doesn't hold (i. e. for which we have opposite sharp inequality), we can decrease the value of σ^2 still fulfilling the constraint given in (9). For $\beta = \hat{\beta}^{(SW,n,w,\rho)}$ we have equality in (12). Hence, also partial derivatives of $\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right)\rho\left(\frac{r^2_{(i)}(\beta)}{\sigma^2}\right)$ with respect to the coordinates of β are equal to 0. It means that the S-weighted estimator has to fulfil the normal equations[2]

$$\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right) X_{j_i}\psi\left(\frac{Y_{j_i} - X'_{j_i}\beta}{\sigma^2}\right) = 0 \tag{13}$$

where j_i is a such index that $r^2_{j_i}(\beta) = r^2_{(i)}(\beta)$. Let us denote

$$\tau(\beta, j_i) = i \in \{1, 2, ..., n\} \qquad \Leftrightarrow \qquad r^2_{j_i}(\beta) = r^2_{(i)}(\beta) \tag{14}$$

(compare Hájek, J., Z. Šidák (1967)). It allows to write the normal equation (13) in a bit more convenient way

$$\sum_{i=1}^{n} w\left(\frac{\tau(\beta, j_i) - 1}{n}\right) X_{j_i}\psi\left(\frac{Y_{j_i} - X'_{j_i}\beta}{\sigma^2}\right)$$

$$= \sum_{j=1}^{n} w\left(\frac{\tau(\beta, j) - 1}{n}\right) X_j\psi\left(\frac{Y_j - X'_j\beta}{\sigma^2}\right) = 0 \tag{15}$$

where $\psi = \rho'$ and of course $\tau(\beta, j)$ again depends on $\omega \in \Omega$. Further, denoting the indicator of a set A by $I\{A\}$, we put

$$F^{(n)}_\beta(r) = \frac{1}{n}\sum_{i=1}^{n} I\{|r_j(\beta)| < r\} = \frac{1}{n}\sum_{i=1}^{n} I\left\{|Y_j - X'_j\beta| < r\right\} \tag{16}$$

and easy verify that (see Víšek (2011a))

$$\frac{\tau(\beta, j) - 1}{n} = F^{(n)}_\beta(|r_j(\beta)|).$$

Finally, we can rewrite the normal equations (15) as

$$\sum_{i=1}^{n} w\left(F^{(n)}_\beta(|r_j(\beta)|)\right) X_j\psi\left(\frac{Y_j - X'_j\beta}{\sigma^2}\right) = 0. \tag{17}$$

[2]These normal equations can be derived also by an alternative way, as for LWS, see Víšek (2011a), but it would take more space.

For $\sigma^2 = 1$ and $\psi(r) = r$, the equations (17) coincide with the *normal equations for the LWS* in Víšek (2011a). They allowed, employing the generalized form of Kolmogor-Smirnov result (see Lemma 1), to prove the consistency of $\hat{\beta}^{(LWS,n,w)}$. We will use it below.

As $\psi(0) = 0$, the normal equation (17) can be written as

$$\sum_{\left\{j\,:\,(Y_j - X_j'\beta)\neq 0\right\}} w\left(F_\beta^{(n)}(|r_j(\beta)|)\right) X_j \psi\left(\frac{Y_j - X_j'\beta}{\sigma^2}\right)$$

$$= \sum_{\left\{j\,:\,(Y_j - X_j'\beta)\neq 0\right\}} w\left(F_\beta^{(n)}(|r_j(\beta)|)\right) X_j \left[\psi\left(\frac{Y_j - X_j'\beta}{\sigma^2}\right) \cdot \frac{\sigma^2}{Y_j - X_j'\beta}\right] \left(Y_j - X_j'\beta\right) = 0. \tag{18}$$

Let us denote

$$v(r, \sigma^2) = \psi(\frac{r}{\sigma^2})\frac{\sigma^2}{r}$$

and notice that then $v(r, \sigma^2)$ is non-negative and symmetric in r, i. e. we have $v(r, \sigma^2) = v(|r|, \sigma^2)$. Finally, recall that $r_j(\beta) = Y_j - X_j'\beta$. Then

$$\sum_{\left\{j\,:\,(Y_j - X_j'\beta)\neq 0\right\}} w\left(F_\beta^{(n)}(|r_j(\beta)|)\right) \cdot v\left(|r_j(\beta)|, \sigma^2\right) X_j \left(Y_j - X_j'\beta\right) = 0. \tag{19}$$

Prior to continuing in considerations, let us recall the basic idea of Hampel's approach to robustification of the classical point estimators. The idea says that it is equivalent whether we have at hand observations or their empirical d. f. (e. d. f.). In other words, that we can consider any point estimator as a function of e. d. f.. In the regression framework it means that any function which depends on the residuals can be written as a function of e. d. f. of residuals, see Hampel et al. (1986). However, applying this idea directly on (19), we would (generally) introduce a discontinuity into the modified version of (19). That is why we utilize instead of $F_\beta^{(n)}(|r_j(\beta)|)$ its continuous modification $\tilde{F}_\beta^{(n)}(|r_j(\beta)|)$, as follows.

For the technical convenience let us write $|r|_j(\beta)$ instead $|r_j(\beta)|$. Then $|r|_{(1)}(\beta), |r|_{(2)}(\beta), ..., |r|_{(n)}(\beta)$ denotes the order statistics of absolute values of residuals. Finally, let us consider a modification of the e. d. f. $F_\beta^{(n)}(v)$, say $\tilde{F}_\beta^{(n)}(v)$ which coincides with $F_\beta^{(n)}(v)$ at $|r_j(\beta)|,\ i = 1, 2, ..., n$ and is continuous and strictly monotone between any pair of $|r|_{(j)}(\beta)$ and $|r|_{(j+1)}(\beta)$. Then we have one-to-one relation between $|r|_j(\beta)$'s and $\tilde{F}_\beta^{(n)}(v)$. Finally, the normal equations (19) can be rewritten as

$$\sum_{\left\{j\,:\,(Y_j - X_j'\beta)\neq 0\right\}} w\left(\tilde{F}_\beta^{(n)}(|r_j(\beta)|)\right) \cdot \tilde{v}\left(\tilde{F}_\beta^{(n)}(|r_j(\beta)|), \sigma^2\right) X_j \left(Y_j - X_j'\beta\right) = 0 \tag{20}$$

where $\tilde{v}(r) : [0, 1] \rightarrow [0, 1]$ is for any value of σ^2 non-negative and non-increasing. Finally, as the product of two non-increasing functions is again non-increasing, we arrive at (notice that we can sum over all indexes, as for i's with $Y_j - X_j'\beta = 0$, the whole summand is also 0)

$$\mathbb{N}E_{Y,X,n}(\beta,\sigma^2) = \sum_{j=1}^{n} \tilde{w}\left(F_\beta^{(n)}(|r_j(\beta)|),\sigma^2\right) X_j\left(Y_j - X_j'\beta\right) = 0 \tag{21}$$

where $\tilde{w}$ fulfils all requirements we have asked for consistency of $\hat{\beta}^{(LWS,n,w)}$ (even under the heteroscedasticity), see Víšek (2011b). Of course, σ^2 is a nuisance parameter which however changes only "rate of decrease" of the function $\tilde{w}$ (i. e. the part of $\tilde{w}$ between h/n and g/n). Now, we are going to cope with this problem.

Due to monotonicity of any d. f. we have can modify Lemma 1 from Víšek (2011b):

Lemma 1. *Let (C1) hold. Recalling that e_i's have different variances σ_i^2, let us denote $F_{i,\beta}(v) = P\left(|Y_i - X_i'\beta| < v\right)$ and put*

$$\overline{F}_{n,\beta}(v) = \frac{1}{n}\sum_{i=1}^{n} F_{i,\beta}(v). \tag{22}$$

Then for any $\varepsilon > 0$ there is a constant K_ε and $n_\varepsilon \in \mathcal{N}$ so that for all $n > n_\varepsilon$

$$P\left(\left\{\omega \in \Omega : \sup_{v\in R^+}\ \sup_{\beta\in \mathbb{R}^p} \sqrt{n}\left|\tilde{F}_\beta^{(n)}(v) - \overline{F}_{n,\beta}(v)\right| < K_\varepsilon\right\}\right) > 1-\varepsilon. \tag{23}$$

This is a key result which allowed in Víšek (2011a) to prove the (weak) consistency of any sequence of solutions of (21) where we fix the value of σ^2. But due to the continuity in σ of (21), we modify the steps by which we proved the consistency of $\hat{\beta}^{(LWS,n,w)}$. Let us briefly recall them, already in a generalized form.

Lemma 2. *Let (C1) and (WR) be fulfilled. Then for any $\varepsilon > 0$ and $\sigma^2 > 0$ there is $\theta > 0$, $\delta > 0$, $\Delta > 0$ and $n_{\varepsilon,\Delta} \in \mathcal{N}$ such that for any $n > n_{\varepsilon,\Delta}$*

$$P\left(\left\{\omega \in \Omega : \inf_{\|\beta\|\geq\theta,\ \tilde{\sigma}^2\in(\sigma^2-\delta,\sigma^2+\delta)} -\frac{1}{n}\beta'\mathbb{N}E_{Y,X,n}(\beta,\tilde{\sigma}^2) > \Delta\right\}\right) > 1-\varepsilon,$$

for the proof see Lemma 2.2 in Víšek (2011b). Similarly, we derive from Lemma 2.3 of Víšek (2011b):

Lemma 3. *Let (C1) and (WR) be fulfilled. Then for any $\varepsilon > 0$, $\tau \in (0,1)$, $\zeta > 0$ and $\sigma^2 > 0$ there is $n_{\varepsilon,\delta,\zeta,\sigma^2} \in \mathcal{N}$ and $\delta > 0$ so that for any $n > n_{\varepsilon,\delta,\zeta,\sigma^2}$ we have*

$$P\left(\left\{\omega \in \Omega : \sup_{\|\beta\|\leq\zeta,\ \tilde{\sigma}^2\in(\sigma^2-\delta,\sigma^2+\delta)} \left|\frac{1}{n}\sum_{i=1}^{n}\left\{\tilde{w}\left(F_\beta^{(n)}(|r_i(\beta)|,\tilde{\sigma}^2)\right)\beta' X_i\left(e_i - X_i'\beta\right)\right.\right.\right.\right.$$
$$\left.\left.\left.\left.-\beta'\mathbb{E}\left[\tilde{w}\left(\overline{F}_{n,\beta}(|r_i(\beta)|,\tilde{\sigma}^2)\right)X_i\left(e_i - X_i'\beta\right)\right]\right\}\right| < \tau\right\}\right) > 1-\varepsilon.$$

Further, without any modification we overtake Lemma 2.4 of Víšek (2011b):

Lemma 4. *Let (C1) hold and moreover* $\frac{1}{n}\sum_{i=1}^{n}\left|1-\sigma_i^2\right| = 0$. *Finally, let* e *be a r.v. distributed according to* $F_e(v)$ *and for any* $\beta \in R^p$ *denote* $F_\beta(v) = P(|e - X_1'\beta| < v)$. *Then for any* $\lambda > 0$

$$\lim_{n\to\infty} \sup_{-\infty<v<\infty} \sup_{\|\beta\|\le\lambda} \left|\overline{F}_{n,\beta}(v) - F_\beta(v)\right| = 0. \tag{24}$$

Finally, we generalize Lemma 2.5 (or Corollary 2.6) of Víšek (2011b):

Lemma 5. *Let (C1) and (WR) be fulfilled. Moreover, let* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}\sigma_i = 1$. *Then for any* $\varepsilon > 0$, $\tau \in (0,1)$ *and* $\zeta > 0$ *there is* $n_{\varepsilon,\delta,\zeta} \in \mathcal{N}$ *so that for any* $n > n_{\varepsilon,\delta,\zeta}$ *we have*

$$P\left(\left\{\omega\in\Omega \;:\; \sup_{\|\beta\|\le\zeta,\ \tilde\sigma^2\in(\sigma^2-\delta,\sigma^2+\delta)} \left|\frac{1}{n}\sum_{i=1}^{n}\tilde w\left(F_\beta^{(n)}(|r_i(\beta)|,\tilde\sigma^2)\right)\beta' X_i\left(e_i - X_i'\beta\right)\right.\right.\right.$$
$$\left.\left.\left. -\beta'\mathbb{E}\left[\tilde w\left(F_\beta(|r(\beta)|,\tilde\sigma^2)\right)X_1\left(e - X_1'\beta\right)\right]\right| < \tau\right\}\right) > 1-\varepsilon.$$

For the final assertion we will need an identification condition.
(C2) *For any fixed* $\sigma^2 > 0$ *there is the only solution of*

$$\mathbb{E}\left[\sum_{i=1}^{n}\tilde w\left(F_\beta^{(n)}(|r_i(\beta)|),\sigma^2\right)X_1\left(e - X_1'\beta\right)\right] = 0.$$

We are not able to prove in full generality - without a long chain of technicalities - the consistency of $\hat\beta^{(SW,n,w,\rho)}$. So we restrict a bit the definition of $\hat\beta^{(SW,n,w,\rho)}$ as follows (compare Víšek (2006a) where we have restricted LTS in a similar way, to be able to give a proof of its consistency):

Definition 4. Let (WR) hold and $C = [a,b]$ be an interval with $0 < a < b < \infty$. Then

$$\hat\beta^{(SWC,n,w,\rho)} = \arg\min_{\beta\in R^p}\left\{\sigma(\beta)\in C \;:\; \frac{1}{n}\sum_{i=1}^{n} w\left(\frac{i-1}{n}\right)\rho\left(\frac{r_{(i)}^2(\beta)}{\sigma^2}\right) = b\right\} \tag{25}$$

where $b = \mathbb{E}\rho(\frac{\varepsilon_1^2}{\sigma_0^2})$, is called the <u>*SC-weighted estimator*</u>.

Repeating the considerations which we made after introducing the S-weighted estimator $\hat\beta^{(SC,n,w,\rho)}$ implies that

$$\hat\beta^{(SWC,n,w,\rho)} = \arg\min_{\beta\in R^p}\left\{\sigma(\beta)\in C : \sum_{i=1}^{n}\tilde w\left(F_\beta^{(n)}(|r_i(\beta)|),\sigma^2\right)X_i\left(Y_i - X_i'\beta\right) = 0\right\}. \tag{26}$$

Theorem 1. *Let* **Conditions** *(C1), (C2) and (WR) be fulfilled. Then any sequence* $\left\{\hat{\beta}^{(SWC,n,w,\rho)}\right\}_{n=1}^{\infty}$ *of the solutions of sequence of normal equations (26) for* $n = 1, 2, ...,$ *is weakly consistent.*

Proof: Applying Lemma 2, ..., 5 we find for any fixed $\sigma^2 \in C$ a non-empty neighborhood, say $N(\sigma^2)$ so that for any $\tilde{\sigma^2} \in N(\sigma^2)$ any sequence of solutions of normal equations

$$I\!N E_{Y,X,n}(\beta, \tilde{\sigma}^2) = \sum_{i=1}^{n} \tilde{w}\left(F_{\beta}^{(n)}(|r_i(\beta)|), \tilde{\sigma}^2\right) X_i \left(Y_i - X_i'\beta\right) = 0 \qquad (27)$$

converges in probability to β^0 and the convergence is uniform over $N(\sigma^2)$. So we create an open cover of the compact set C. Then selecting a cover with finite elements, say $\left\{N(\sigma_k^2)\right\}_{k=1}^{K}$ we apply the usual steps of "merging" the convergence in probability of the estimators $\hat{\beta}^{(SWC,n,w,\rho)}$ in the individual sets $N(\sigma_k^2), k = 1, 2, ..., K$.

4 CONCLUSION

The paper proposed a new estimator - S-weighted estimator - which "covers with roof" both the *least weighted squares* as well a *S-estimator* (special cases of which are also the least median of squares and the least trimmed squares) and inherits the plausible properties of all of them. The consistency is proved for a bit less general version of estimator which is restricted on the case when the variance of disturbances has value from a compact set. This restriction is not substantial from the applicability point of view. On the other hand, at the expense of technically more complicated proofs (with a long chains of approximations) this restriction can be released. The computation can be carry out by a slightly modified algorithm earlier employed for computation of the least weighted squares (or the least trimmed squares), see Víšek (1990) or (1994). The possibility to select the weight function w as well as the objective function ρ from a wide spectrum of functions makes the S-weighted estimator very flexible. Finally, the speed of today IT means allow to tailor both of them - applying some kind of *forward search* - just for data in question. The software written in MATLAB is available on the request.

References

1. Atkinson, A. C., M. Riani (2000): *Robust Diagnostic Regresion Analysis.* New York: Springer Verlag.
2. Hájek, J., Z. Šidák (1967): *Theory of Rank Test.* Academic Press, New York.
3. Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel (1986): *Robust Statistics – The Approach Based on Influence Functions.* New York: J.Wiley & Sons.
4. Rousseeuw, P.J. (1984): Least median of square regression. *Journal of Amer. Statist. Association 79, pp. 871-880.*

5. Rousseeuw, P. J., V. Yohai (1984): Robust regressiom by means of S-estimators. In: *Robust and Nonlinear Time Series Analysis. eds. J. Franke, W. Hårdle and R. D. Martin, Lecture Notes in Statistics No. 26 Springer Verlag, New York, 256-272.*
6. Siegel, A. F. (1982): Robust regression using repeated medians. *Biometrica, 69, 242 - 244.*
7. Víšek, J. Á. (1990): Empirical study of estimators of coefficients of linear regression model. *Technical report of Institute of Information Theory and Automation, Czechoslovak Academy of Sciences (1990), number 1699.*
8. Víšek, J. Á. (1994b): A cautionary note on the method of Least Median of Squares reconsidered, *Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, Lachout, P., Víšek, J.Á. (eds), Academy of Sciences of the Czech Republic, Prague, 1994, 254 - 259.*
9. Víšek, J. Á. (2000): Regression with high breakdown point. *Robust 2000 (eds. Jaromír Antoch & Gejza Dohnal, published by Union of Czech Mathematicians and Physicists), 2001, 324 - 356.*
10. Víšek, J. Á. (2006a): The least trimmed squares. Part I - Consistency. *Kybernetika (2006), vol 42, 1 - 36.*
11. Víšek, J. Á. (2006b): Instrumental Weighted Variables - algorithm. *Proceedings of the COMPSTAT 2006, eds. A. Rizzi & M. Vichi, Physica-Verlag (Springer Company) Heidelberg, 777-786.*
12. Víšek, J. Á. (2011a): Consistency of the least weighted squares under heteroscedasticity. *Kybernetika 47 , 179-206.*
13. Víšek, J. Á. (2011b): Empirical distribution function under heteroscedasticity. *Statistics 45, 497-508.*