Department of Mathematics
Faculty of Mechanical Engineering
Slovak University of Technology in Bratislava
4th International Conference
APLIMAT 2005

TRANSFORMATION AND PROBABILITY MODELS

VÁVRA František (ČR), NOVÝ Pavel (ČR), NETRVALOVÁ Arnoštka (ČR), NEUMANOVÁ Martina (ČR), VOKÁČOVÁ Kateřina (ČR)

Abstract. The proposed paper deals with one of methods for study and modelling the relation of two probability descriptions (estimation of distribution and its theoretical model is one example of using) at distribution functions level. The essence of the used concept is a transformation of random variables. A version of duality between transformation (distortion) of distribution function and transformation of random variables is investigated too. The sensitivity of such descriptions is analyzed. Some algorithmic techniques are introduced (from statistical applications point of view). Paper extends our presentation [1].

Key words. Transformation of random variables, distortion, divergence, distribution

1 Method and model

We consider all of used cumulative distributions to be continuous and invertible. Further we consider a random variable η defined as an invertible (increasing) transformation $t()$ of a random variable ξ . Let $\eta = t(\xi)$. We compute the representations of the distributions η in terms of the distributions of ξ. By the definition of the cumulative distribution F_η we have:

$$F_\eta(x) = P\{\eta < x\} = P\{t(\xi) < x\} = P\{\xi < t^{-1}(x)\} = F_\xi(t^{-1}(x)), \tag{1}$$

where notation $t^{-1}()$ represents the inversion of the transformation $t()$.

Substitution $y = t^{-1}(x) \Leftrightarrow x = t(y)$ into (1) we obtain:

$$F_\eta(x) = F_\xi(y) \Rightarrow F_\eta(t(y)) = F_\xi(y) \Rightarrow \quad t(x) = F_\eta^{-1}(F_\xi(x)), \tag{2}$$

$t()$ is continuous and increasing on the set $\{x : 0 < F_\xi(x) < 1\}$.

In similar way we compute the distributions η in terms of a distortion [1, 2].

$$F_\eta(x) = G(F_\xi(x)) \tag{3}$$

The use of substitution $y = F_\xi(x) \Leftrightarrow x = F_\xi^{-1}(y)$ means

$$G(x) = F_\eta\,(F_\xi^{-1}(x)) \tag{4}$$

If we take into account the characteristics of cumulative distribution functions, it is clear that $G(x)$ is continuous and increasing function from $\langle 0,1\rangle$ to $\langle 0,1\rangle$. The transformation and distortion – (2) and (4) – are equivalent in the case, where both $F_\eta(), F_\xi()$ are well known. The properties distortion functions were studied in [1]. In this paper we will demonstrate some aspects of a transformation view.

2 Case $F_\eta()$ is good known and $F_\xi()$ is estimated

This modification represents e.g. situation, where $F_\eta()$ represents a model and $F_\xi()$ is empirical distribution function[1]. Consider that $t(x)=F_\eta^{-1}(F_\xi(x))$ is impossible to obtain without errors. Thus it's useful define

$$t_{e(x)}(x)=F_\eta^{-1}(F_\xi(x)+e(x)), \tag{5}$$

where $t_{e(x)}(x)$ is a real model of obtained transformation. Because $0\le F_\xi(x)+e(x)\le 1$, $e(x)$ fits $-F_\xi(x)\le e(x)\le 1-F_\xi(x)$. Now, it's possible to define the sensitivity functional (shorter sensitivity) as:

$$s(t(x))=\lim_{e(x)\to 0}\frac{t_{e(x)}(x)-t(x)}{e(x)}. \tag{6}$$

In (6) $e(x)\to 0; e(x)\ne 0$ stands for uniform convergence to zero function on the definition's domain $F_\xi(x)$. This sensitivity definition is derived (and motivated) from Taylor's expansion $t_{e(x)}(x)\cong t(x)+s(t(x))e(x)$. Now it's obvious when $|s(t(x))|<1$ then influence of the error $e(x)$ is reduced and in case $|s(t(x))|>1$ is enlarged. From the practical point of view it's interesting to study domain $|s(t(x))|<1$, now. Using (6) and l' Hospital's rule is obtained:

$$s(t(x))=\frac{1}{f_\eta(t(x))}, \tag{7}$$

where $f_\eta(x)=\frac{d}{dx}F_\eta(x)$ is the probability density function of the distribution $F_\eta(x)$.

From (2) and (7) we have $s(t(x))=\dfrac{1}{f_\eta(F_\eta^{-1}(F_\xi(x)))}$. (8)

The error reducing condition $|s(t(x))|<1$ for domain implies $f_\eta(F_\eta^{-1}(F_\xi(x)))>1$, because of non negativity $f_\eta()$. It's necessary to remind that the condition $|s(t(x))|<1$ is sufficient for any pair $F_\eta(), F_\xi()$ (not only in case presented in this paper). Each computation leads to numerical representation with some errors.

---

[1] By the initial assumptions on distribution functions, empirical distribution function is defined in this way: Let $x_1,\dots,x_n$ be a random sample and $x_{(1)},\dots,x_{(n)}$ its ordered version. For any $x_{(i)}$ is $F_{empir.}(x_{(i)})=\frac{i}{n}$; for $x\le x_{(1)}\Rightarrow F_{empir.}(x)=0$; for $x\ge x_{(n)}\Rightarrow F_{empir.}(x)=1$ and for any x between $x_{(i)}$ and $x_{(i+1)}$ $i<n$ $F_{empir.}(x)$ is line from $(x_{(i)},F_{empir.}(x_{(i)}))$ to $(x_{(i+1)},F_{empir.}(x_{(i+1)}))$. Then $F_{empir.}(x)$ is continuous (on real axis R_1) and increasing on the set $\{x: 0<F_{empir.}(x)<1\}$.

3 Example: PX50 log-return

Data of log-return Prague stock exchange PX50 index are used (time segment from 7.9.1993 to 15.9.2005).

Log-return $r(t)$ is defined as log-ratio $r(t) = log_e\left(\frac{i(t)}{i(t-1)}\right)$, where $i(t)$ is daily PX50 value in the $t-day$. As $F_\eta(x)$ we use normal approximation with the empirical mean and empirical variance. And $F_\xi()$ is empirical distribution function of the log-return PX50. Both $F_\eta(), F_\xi()$ are demonstrated in Figure 1.

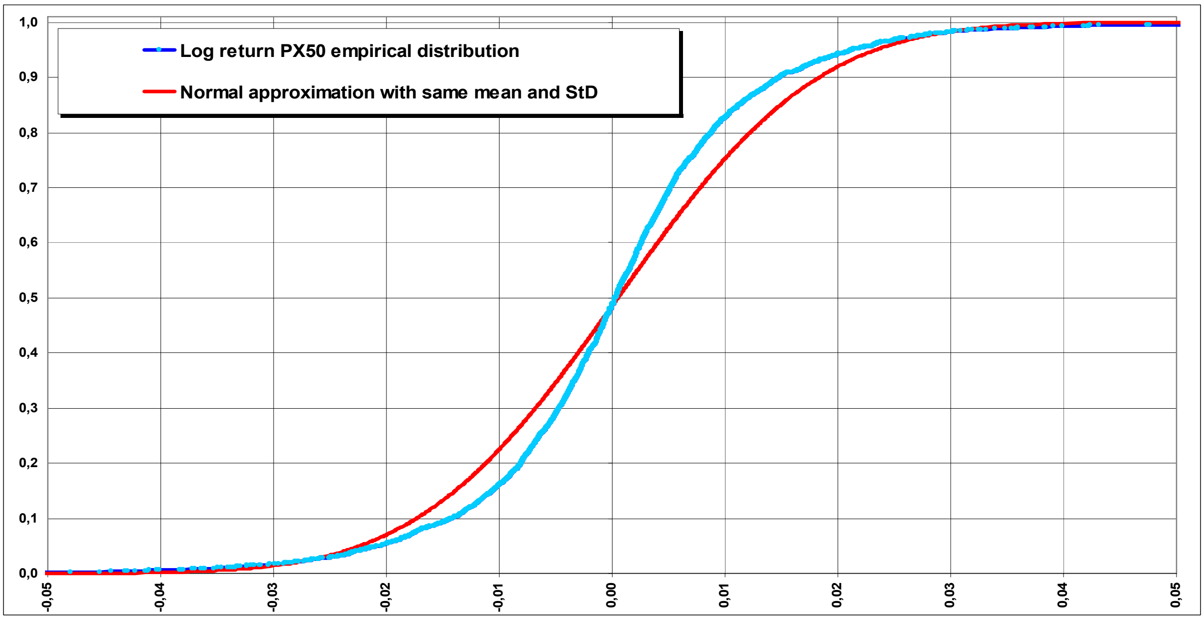

Figure 1 Log-return PX50 empirical distribution and its normal approximation.

Transformation (2) and sensitivity (7) are plotted in Figure 2.

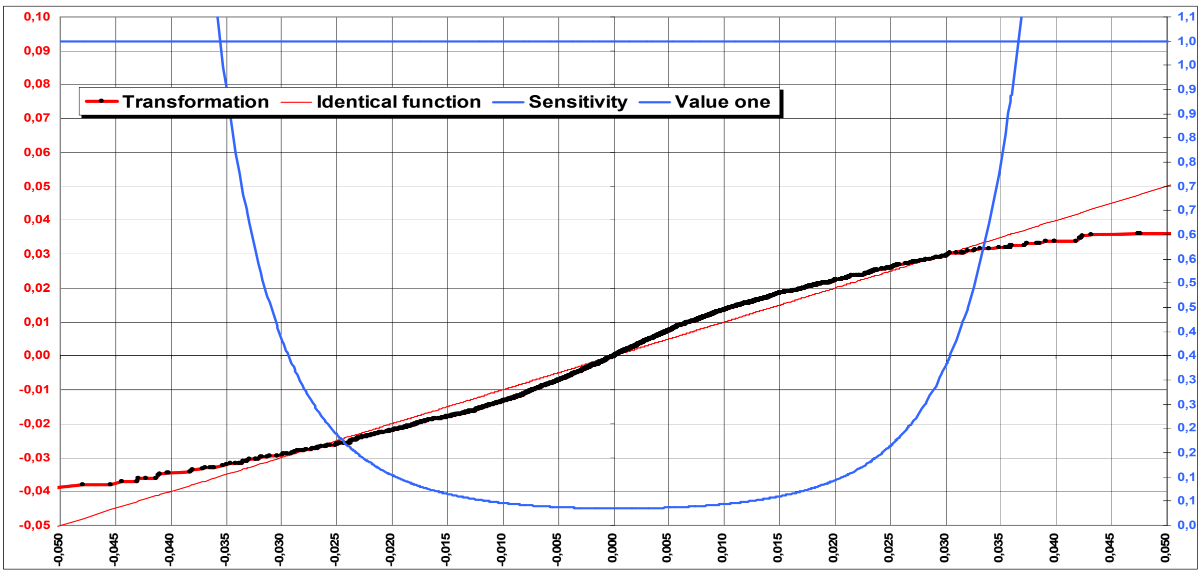

Figure 2 Transformation and sensitivity. Identity and transformation are on the left axis and sensitivity and one are on the right.

From Figure 2 is visible that the error contracting area is (approximately) interval from -0.035 to +0.035. It represents interval from 0.9656 to 1. 0356 in classical PX50 return $\left(\frac{i(t)}{i(t-1)}\right)$. The distortion $G(x)$ is shown in Figure 3, for the completeness.

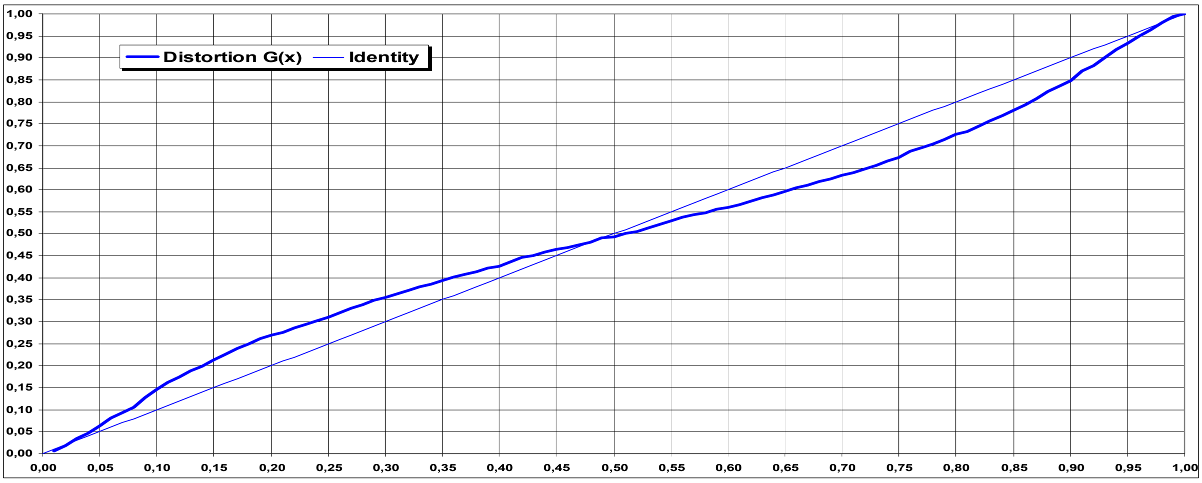

Figure 3 Distortion of the pair $F_\eta(), F_\xi()$. Normal approximation log-return PX50 and its empirical distribution.

4 Example: ČEZ stock return

Data of the return Prague stock exchange CEZ-stock are used (time segment from 1.12.2000 to 31.8.2005).

Return $d(t)$ is defined as ratio $d(t)=\left(\frac{i(t)}{i(t-1)}\right)$, where $i(t)$ is daily stock value in the $t-day$.

As $F_\eta(x)$ we use the normal approximation with the empirical mean and empirical variance. And $F_\xi()$ is empirical distribution function of the daily stock return is. Both $F_\eta(), F_\xi()$ are demonstrated in Figure 4.

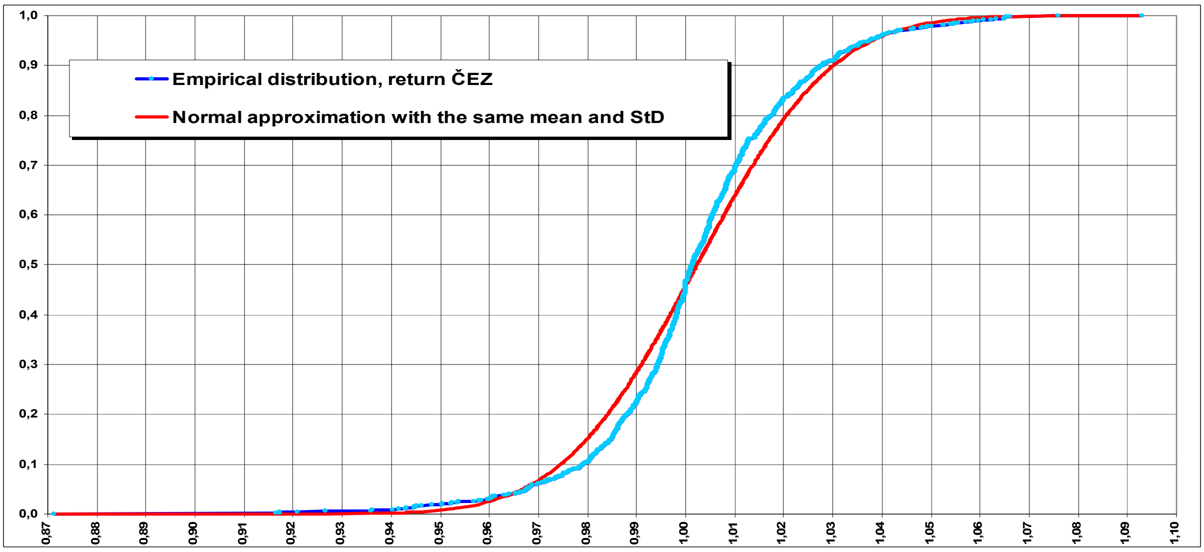

Figure 4 Return CEZ stock empirical distribution and its normal approximation.

Transformation (2) and sensitivity (7), on this example are visible in Figure 5.

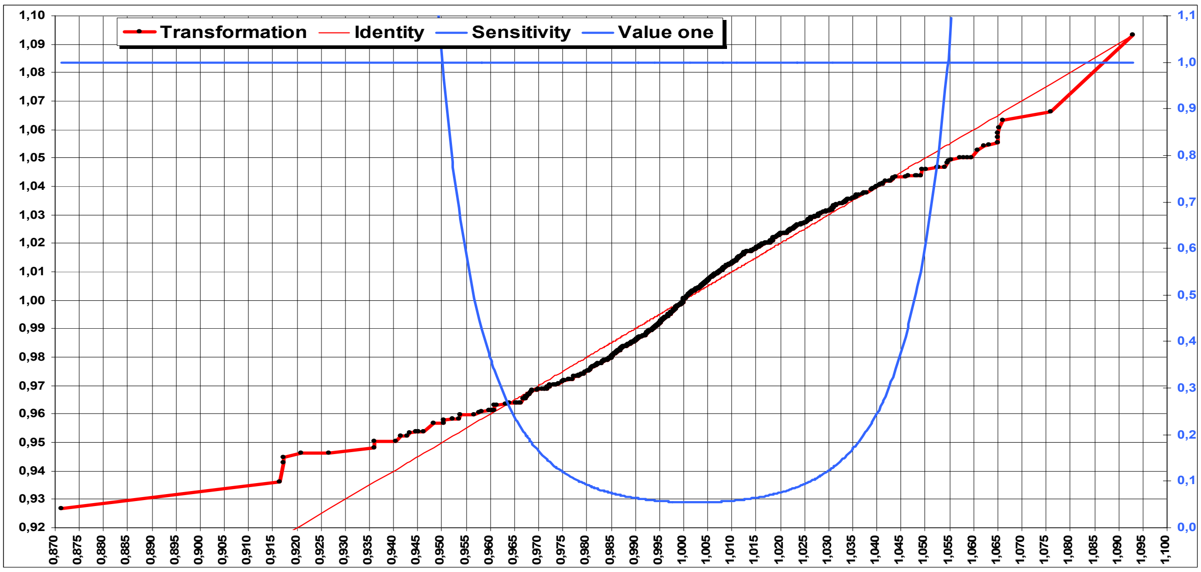

Figure 5 Transformation and sensitivity. Identity and transformation are on the left axis and sensitivity and value one are on the right (example CEZ stock).

From Figure 5 is clear that the error contracting area is (approximately) interval from 0.95 to +1.05.

5 Example: Cauchy random variable simulation

This case may represent situation, where $F_\eta()$ describes a model and $F_\xi()$ is empirical distribution of one thousand simulated samples of Cauchy random variable (location parameter=0, scale parameter=1). As the $F_\eta(x)$ we use normal approximation with the empirical mean and empirical variance (but mean and variance in this case don't exists). And $F_\xi()$ is empirical distribution function of the simulated data. Both $F_\eta(), F_\xi()$ are demonstrated in Figure 6.

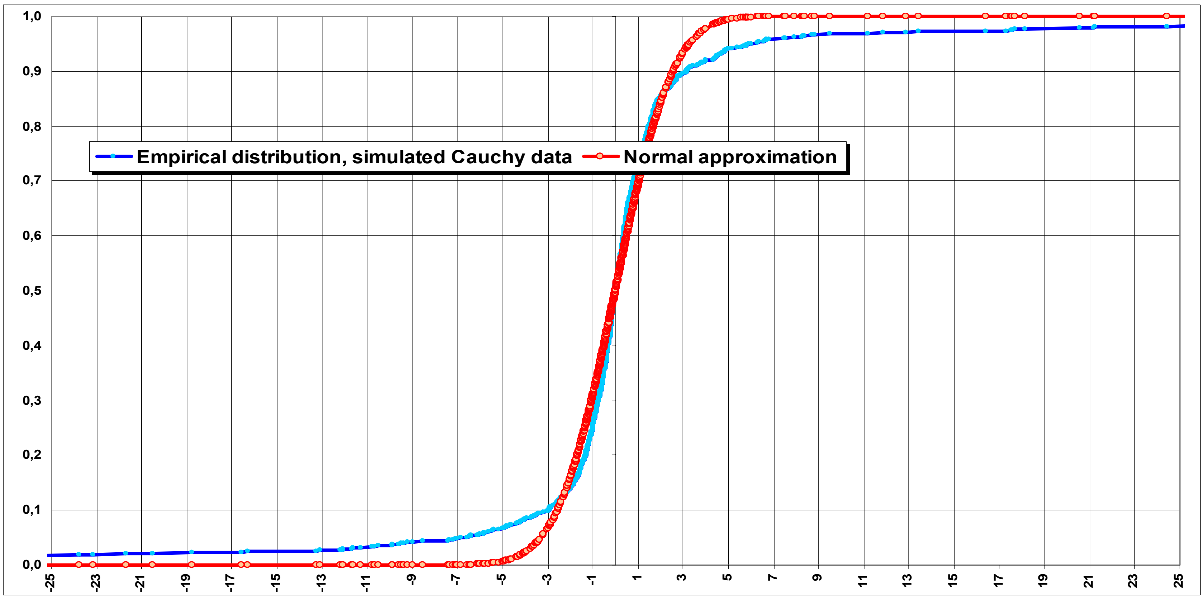

Figure 6 Simulated Cauchy data empirical distribution and its normal approximation.

Transformation (2) and sensitivity (7), for this example are in Figure 7.

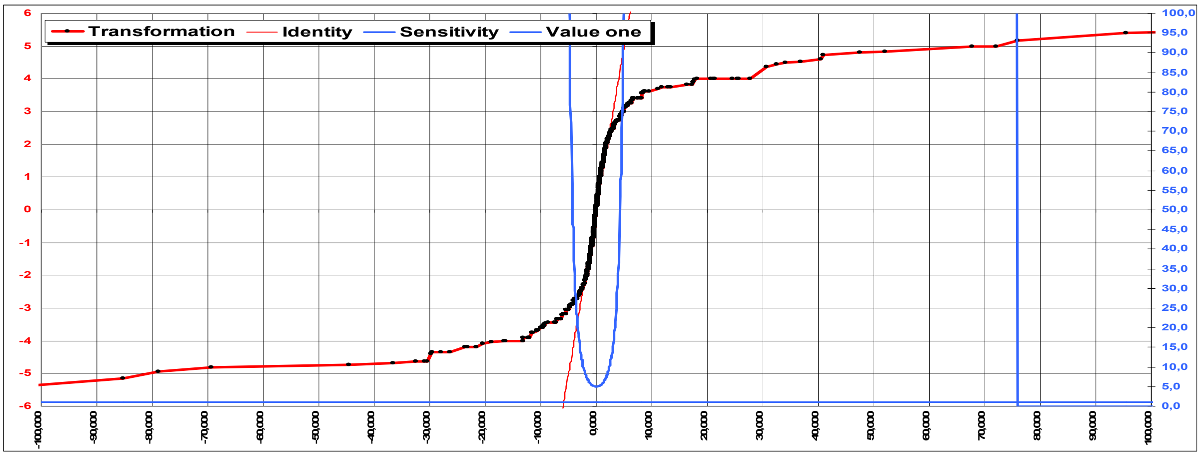

Figure 7 Transformation and sensitivity. Identity and transformation are on the left axis and sensitivity and value one are on the right (example Cauchy data).

From the Figure 7 is clear that the error contracting area does not exist.

6 Conclusions

Considered transformation is very useful, but very dangerous as well. If the partition (random variable definition area) into two sub-areas is not analyzed, then generating large errors is possible. In the case of distribution with fat tails possibility converts to certainty. The error contracting domain is a good description for accepting area of transformation model and its complement is transformation rejecting area. This paper shows that the error contracting area (and complementary transformation rejecting area) depends on a choice $F_\eta()$, too. If the fat tails occurred, there is better to use switching models [3].

References

[1] Vávra, F., Nový, P., Mašková, H., Netrvalová, A.: *Distortion of probability models*, 4th International Conference APLIMAT 2005, Slovak University of technology, Bratislava.

[2] Reesor, R. M., McLeish, Don L.: *Risk, Entropy, and the Transformation of Distributions*, Bank of Canada, Working Paper 2002-11.

[3] Vojtíšková, M.: *Asset price forecasting – regular and singular factors influence,* 5. mezinárodní konference Finanční řízení podniku a finančních institucí, VŠB-TU Ostrava, Ekonomická fakulta, katedra Financí, Ostrava, 2005

Contact address

Doc. Ing. František Vávra, CSc., vavra@kiv.zcu.cz
Ing. Pavel Nový, Ph. D., novyp@kiv.zcu.cz
Ing. Arnoštka Netrvalová, netrvalo@kiv.zcu.cz
Ing. Martina Neumanová, mneumano@kiv.zcu.cz
Ing. Kateřina Vokáčová, katerina.vokacova@cca.cz

Department of Computer Science and Engineering, Faculty of Applied Sciences
University of West Bohemia, Univerzitní 22, 306 14 Plzeň, Czech Republic