

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Kateřina Podolská

Grafické modely v analýze spojitých finančních dat

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jitka Zichová, Dr.

Studijní program: Matematika, studijní obor finanční matematika

2006

Chtěla bych poděkovat zejména vedoucí práce RNDr. Jitce Zichové, Dr. za nevšedně vstřícný přístup a cenné připomínky k práci. Dále děkuji tvůrci programu Backward1 [4]. Také velice děkuji mému manželovi za pomoc při zvládnutí technologie L^AT_EX, i našim dcerám za trpělivost během psaní bakalářské práce.

Prohlašuji, že jsem svou bakalářskou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 14.května 2006

Kateřina Podolská

Obsah

Úvod	4
1 Základní pojmy	5
1.1 Grafy podmíněných nezávislostí	5
1.2 Interpretace grafu	10
2 Věrohodnostní přístup	13
2.1 Zvolená metoda	13
2.2 Testování vstupních dat	14
2.3 Selekcční algoritmus	14
3 Zpracování dat a interpretace výsledků	16
3.1 Postup přípravy dat ke zpracování	16
3.2 Výsledky zpracování	17
4 Závěr	22
4.1 Interpretace výsledků	22
4.2 Poznámky k výpočtu denních kurzů ČNB	22
Literatura	24
Příloha:	
Protokoly zpracování programem Backward1 (1999–2005)	25

Název práce: Grafické modely v analýze spojitých finančních dat

Autor: Kateřina Podolská

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Jitka Zichová, Dr.

e-mail vedoucího: Jitka.Zichova@mff.cuni.cz

Abstrakt: Hlavním cílem této bakalářské práce je aplikovat grafické modely určené pro zpracování dat se spojitým rozdělením na data z finanční praxe. Pro tento účel byly zvoleny denní měnové kurzy z databáze ČNB. Pro výběr konkrétního grafického modelu byl použit program Backward1, vytvořený jako součást diplomové práce [4], realizovaný v systému Mathematica 4.0.

Klíčová slova: grafické modely, analýza spojitých finančních dat

Title: Graphical Models in Continuous Financial Data Analysis

Author: Kateřina Podolská

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jitka Zichová, Dr.

Supervisor's e-mail address: Jitka.Zichova@mff.cuni.cz

Abstract: The main topic of this bachelor thesis is the application of graphical models developed for data processing of the continuous distribution on finance data. For this purpose, daily exchange rates from the ČNB database were selected. The program Backward1, which was created as a part of the thesis [4] under the system Mathematica 4.0, was used for selection of the real graphical model.

Keywords: Graphical Models, Continuous Data, Financial Analysis

Úvod

Důležitým nástrojem statistické analýzy mnohorozměrných dat se v poslední době stávají grafické modely. Umožňují studium struktury závislosti více proměnných. Uplatňují se proto stále častěji ve finanční analýze.

Hlavním cílem této bakalářské práce byla aplikace metod a algoritmu vytvořených v diplomové práci [4] pro analýzu spojitých dat z finanční praxe. Jako tato data byla zvolena databáze měnových kurzů ČNB.

První kapitola popisuje formou stručného přehledu základní pojmy z teorie grafů a mnohorozměrné statistiky. Věnuje se také interpretaci grafů pro dvou, tří a čtyřsložkové náhodné vektory. Parametry těchto grafů jsou uvedeny v přehledové tabulce.

Ve druhé kapitole je popsána zvolená metoda, postup testování zpracovávaných dat a selekční algoritmus pro výběr grafického modelu s dobrou shodou s daty. Je zde také stručný popis programu Backward1, který byl pro výběr modelu prakticky použit.

Třetí kapitola se věnuje postupu přípravy dat ke zpracování a dále pak samotnému zpracování dat. Je zde uveden také popis výsledků a jejich interpretace.

Kapitola 1

Základní pojmy

1.1 Grafy podmíněných nezávislostí

Označme $P_X(x)$ pravděpodobnost, že diskrétní náhodná veličina X nabývá hodnoty x .

Podmíněnou pravděpodobnost jevu $X = x$ při dané hodnotě y náhodné veličiny Y značme $P_{X|Y}(x; y)$.

Má-li náhodná veličina X spojité rozdělení, její hustotu značme f_X .

Podmíněnou hustotu veličiny X v bodě x při dané hodnotě y náhodné veličiny Y zapisujeme $f_{X|Y}(x; y)$. Pro náhodný vektor (X, Y) budeme používat značení P_{XY} a f_{XY} místo $P_{(X,Y)}$ a $f_{(X,Y)}$.

Střední hodnotu náhodné veličiny X značíme $E(X)$, dále zapisujeme EX .

Kovariance náhodných veličin X a Y je definována vztahem

$$\text{cov}(X, Y) = E(X - EX)(Y - EY)$$

Rozptyl je definován vztahem

$$\text{var}(X) = \text{cov}(X, X)$$

Korelaci náhodných veličin X a Y s nenulovým rozptylem definujeme jako

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}$$

Náhodný vektor je vektor

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$$

sestavený z p náhodných veličin (X_1, X_2, \dots, X_p) .

Střední hodnota náhodného vektoru X je vektor $E(X)$, složený ze středních hodnot složek X .

Varianční matice V náhodného vektoru X je matice $\text{var}(X)$, je symetrická a pozitivně semidefinitní. Dále předpokládáme, že $\text{var}(X)$ je pozitivně definitní, tedy existuje její **inverze $D = \text{var}(X)^{-1}$** .

Parciální kovarianci Y a Z při daném X definujeme vztahem

$$\text{cov}(Y, Z|X) = \text{cov}(Y - \hat{Y}(X), Z - \hat{Z}(X))$$

$\hat{Y}(X)$... lineární odhad metodou nejmenších čtverců Y pomocí X ,

$$\hat{Y}(X) = \text{cov}(Y, X)\text{var}(X)^{-1}X$$

$\hat{Z}(X)$... lineární odhad metodou nejmenších čtverců Z pomocí X ,

$$\hat{Z}(X) = \text{cov}(Z, X)\text{var}(X)^{-1}X$$

Parciální rozptyl Y při daném X definujeme vztahem

$$\text{var}(Y|X) = \text{cov}(Y, Y|X)$$

Koeficient parciální korelace mezi náhodnými veličinami Y a Z při daném X je dána vztahem

$$\text{corr}(Y, Z|X) = \frac{\text{cov}(Y, Z|X)}{\sqrt{\text{var}(Y|X) \text{var}(Z|X)}}$$

Nezávislé jsou náhodné vektory X a Y , spojitě rozdělené, právě tehdy, když jejich sdružená hustota f_{XY} splňuje rovnost

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

pro všechna x a y . Tento vztah bude značen $X \perp Y$.

Nezávislé při dané hodnotě vektoru X jsou náhodné vektory Y a Z , spojitě rozdělené, právě když podmíněná hustota $f_{YZ|X}$ splňuje rovnost

$$f_{YZ|X}(y, z; x) = f_{Y|X}(y; x)f_{Z|X}(z; x)$$

pro všechna y a z a pro všechna x , pro která $f_X(x) > 0$. Tento vztah nazýváme podmíněná nezávislost. Budeme ho zapisovat $Y \perp Z | X$.

Graf G je uspořádaná dvojice množin (V, E) .

Vrcholy nazýváme prvky množiny V .

Hrany nazýváme prvky množiny E tvořené dvouprvkovými podmnožinami množiny V .

Je-li $i, j \in E$, říkáme, že v grafu G je **hrana mezi vrcholy i a j** . Vrcholy i a j jsou **spojené**, je-li v grafu G hrana mezi vrcholy i a j .

Cesta spojující vrcholy i a j délky m je posloupnost vrcholů i_1, i_2, \dots, i_m taková, že $i_1 = i$ a $i_m = j$ anebo $i_1 = j$ a $i_m = i$, a že pro každé $l = 1, 2, \dots, m - 1$ platí $i_l, i_{l+1} \in E$. Je-li navíc $i = j$, potom posloupnost $l = 1, 2, \dots, m - 1$ nazýváme cyklus délky $m - 1$.

Nechť posloupnost i_1, i_2, \dots, i_m vrcholů grafu G je cyklus délky $m \geq 4$. Tento cyklus obsahuje **chordu**, jestliže existují indexy $j, k \in 1, 2, \dots, m$, pro které platí $j - k > 1$ a $(j, k) \neq (m, 1)$, takové, že vrcholy i_j a i_k jsou spojené.

Graf G je triangulovaný, jestliže každý cyklus v G délky $m \geq 4$ obsahuje chordu.

Hranice množiny $a \subseteq V$ je množina vrcholů z V , které nejsou v a , a jsou spojené vrcholem z a . Značme je $bd(a)$.

Pro označování hranice jednoprvkové množiny obsahující prvek i použijeme symbol $bd(i)$ namísto $bd(\{i\})$.

Podmnožina vrcholů $a \subseteq V$ **separuje vrcholy i a j** , jestliže každá cesta spojující vrcholy i a j obsahuje alespoň jeden vrchol z a . Množina a odděluje množiny b a c , kde $b \subseteq V$, $c \subseteq V$, jestliže odděluje každé dva vrcholy i a j , $i \in b$ a $j \in c$.

Podgraf grafu G indukovaný množinou $a \subseteq V$ je graf $G_a = (a, E \cap C_a)$, kde C_a je množina obsahující všechny dvouprvkové podmnožiny a .

Úplný je graf nebo indukovaný podgraf, jestliže každý z jeho vrcholů je spojený se všemi jeho ostatními vrcholy.

Klika je taková podmnožina a množiny V , která indukuje úplný podgraf, a která po přidání libovolného dalšího vrcholu z V indukuje podgraf, který

není úplný.

Říkáme, že klika je maximální úplný podgraf. Úplný graf má tedy jedinou kliku, a to celou množinu vrcholů V .

Rozklad grafu G tvoří dvojice (a, b) , $a \subseteq V, b \subseteq V$, jestliže zároveň platí:

- i. $V = a \cup b$,
- ii. Podgraf grafu G indukovaný množinou $a \cap b$ je úplný,
- iii. $a \cap b$ odděluje množiny a a b .

Jsou-li navíc a i b vlastní podmnožiny V , potom dvojice (a, b) tvoří vlastní rozklad grafu G .

Perfektním nazveme uspořádání (C_1, \dots, C_n) všech klik grafu G , pokud pro všechna $i = 2, \dots, n$ platí, že vrcholy z C_i , které jsou obsaženy v některé z předchozích klik, jsou všechny obsaženy v jedné z předchozích klik, tj.

$$\forall i \in \{2, \dots, n\} \exists h \in \{1, \dots, i-1\} : \left(C_i \cap \bigcup_{j=1}^{i-1} C_j \right) \subseteq C_h.$$

Separátory se nazývají množiny

$$S_i = C_i \cap \bigcup_{j=1}^{i-1} C_j, \quad i = 2, \dots, n.$$

Rozložitelným nazveme graf G , jestliže existuje perfektní uspořádání jeho klik.

Graf podmíněných nezávislostí náhodného vektoru X je graf $G = (V, E)$, kde $V = \{1, 2, \dots, k\}$. Hrana $\{i, j\}$ není v E , právě když $X_i \perp X_j \mid X_{K \setminus \{i, j\}}$.

Matice souslednosti grafu \mathbf{G} , $G = (V, E)$, $V = (v_0, \dots, v_n)$, je matice $A_G = (a_{i,j})_{i,j=1}^n$ definovaná předpisem:

$$\begin{aligned} a_{i,j} &= 1 \text{ pro } \{v_i, v_j\} \in E, \\ a_{i,j} &= 0 \text{ jinak.} \end{aligned}$$

Matice souslednosti je čtvercová a symetrická, prvky na diagonále jsou rovny 0. Pokud dva vrcholy nejsou spojeny hranou, prvkem je 0. Pokud jsou spojeny hranou, prvek je roven 1.

Grafický model s grafem $G = (V, E)$ o k vrcholech pro k -rozměrný náhodný vektor $X = X_1, X_2, \dots, X_k$ je skupina pravděpodobnostních rozdělení vektoru X , která splňují podmíněné nezávislosti dané vektorem G .

Pro náhodný vektor $X = X_a, X_b, X_c$ v grafickém modelu platí, že podmnožina vrcholů a **separuje** podmnožiny vrcholů b a c když platí: $X_b \perp X_c \mid X_a$.

Saturovaný model je grafický model určený úplným grafem.

Maximálně věrohodné odhady D a V v grafickém modelu s grafem G založené na náhodném výběru z mnohorozměrného normálního rozdělení označíme jako \hat{D} a \hat{V} . Splňují tyto věrohodnostní rovnice:

$$\begin{aligned} \hat{d}_{ij} &= 0, \text{ pro každé dva vrcholy } i, j \text{ grafu } G, \text{ které nejsou spojené,} \\ \hat{V}_{aa} &= S_{aa}, \text{ pro každou množinu } a \subseteq K, \text{ která je klika,} \end{aligned}$$

kde S_{aa} je submatice výběrové varianční matice S .

Odhady \hat{D} a \hat{V} splňují rovnost $\hat{D} = \hat{V}^{-1}$ a jsou s pravděpodobností 1 určeny jednoznačně.

Vícerozměrné normální rozdělení n -rozměrný náhodný vektor X má n -rozměrné normální rozdělení, jestliže jeho hustota je ve tvaru

$$f_X(x) = (2\pi)^{-\frac{k}{2}} (\det D)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2}(x - \mu)^T D(x - \mu) \right\}$$

χ^2 -rozdělení Náhodná veličina Y má χ^2 -rozdělení s n stupni volnosti, jestliže $Y = X_1^2 + X_2^2 + \dots + X_n^2$, kde X_1, X_2, \dots, X_n jsou nezávislé náhodné veličiny s rozdělením $N(0, 1)$.

Devianci grafu G definujeme jako:

$$dev^f = 2[l(S) - l(\hat{V})]$$

S ... maximálně věrohodný odhad v saturovaném modelu s úplným grafem G_0 , což je výběrová varianční matice.

Graf G , ve kterém chybí f hran, je podgraf grafu G_0 .

\hat{V} ... je maximálně věrohodný odhad v modelu s grafem G .

l ... je logaritmická věrohodnostní funkce.

Deviance dev^f má asymptoticky rozdělení χ_f^2 s f stupni volnosti, kde f je dáno počtem chybějících hran v grafu G .

Je-li G_0 úplný graf, graf G_1 jeho podgraf s f_1 chybějícími hranami a deviancí dev^{f_1} , a graf G_2 podgraf grafu G_1 ($G_0 \supset G_1 \supset G_2$), ve kterém chybí f_2 hran a má devianci dev^{f_2} , pak:

Diferenci deviancí modelů s grafy $G_1 \supset G_2$ definujeme vztahem:

$$dev^* = -[dev^{f_1} - dev^{f_2}].$$

Pokud v grafu G_2 chybí oproti G_1 pouze jedna hrana $\{i, j\}$ ($G_2 = G_1 \setminus \{i, j\}$), nazýváme diferenci jejich deviancí **deviancí vynechané hrany**.

Značíme ji dev_{ij}^* .

1.2 Interpretace grafu

V následující části se budeme zabývat interpretací grafických modelů pro různé náhodné vektory. Jde tedy o modely přípustné pro vystižení struktury vzájemné závislosti složek náhodného vektoru X . Shodu testovaného grafického modelu s daty zamítáme, pokud číselná hodnota deviance překročí příslušnou kritickou hodnotu χ^2 -rozdělení.

Podrobně je problematika uvedena v diplomové práci [5] na str. 23–36. V ní jsou odvozeny maximálně věrohodné odhady prvků varianční matice V . Uvedeme tedy jen stručný přehled interpretace grafů pro dvou, tří a čtyřsložkové náhodné vektory.

Nejsou zde uvedeny modely s úplnými grafy, tedy grafy bez chybějících hran, pro které je $\hat{V} = S$. V takovém případě je deviance nulová. Výsledek je pak interpretován jako model bez chybějících hran, tedy výchozí.

Výčet klik v následujících tabulkách se týká vždy jednoho zástupce dané třídy modelů. Tento zástupce je pak zobrazen v přehledu grafů.

1) **Dvousložkové náhodné vektory** – celkový počet modelů 1

vynechané hrany	poč. modelů dané třídy	počet klik	výčet klik	f	nezávislé složky
1	1	2	{1}{2}	1	{X ₁ }{X ₂ }

2) **Třísložkové náhodné vektory** – celkový počet modelů 7

vynechané hrany	poč. modelů dané třídy	počet klik	výčet klik	f	nezávislé složky
1	3	2	{1, 2}{2, 3}	1	
2	3	2	{1, 2}{3}	2	{X ₁ , X ₂ }{X ₃ }
3	1	3	{1}{2}{3}	3	{X ₁ }{X ₂ }{X ₃ }

3) **Čtyřsložkové náhodné vektory** – celkový počet modelů 63

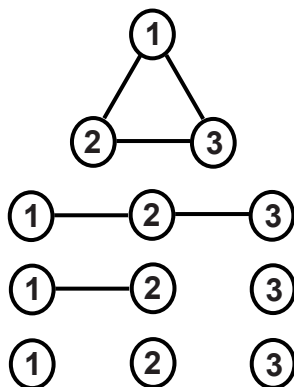
vynechané hrany	poč. modelů dané třídy	počet klik	výčet klik	f	nezávislé složky
1	6	2	{1, 2, 3}{2, 3, 4}	1	
2	3	4	{1, 2}{1, 3}{2, 4}{3, 4}	2	
2	12	2	{1, 2, 3}{3, 4}	2	
3	4	2	{1, 2, 3}{4}	3	{X ₁ , X ₂ , X ₃ }{X ₄ }
3	4	3	{1, 2}{1, 3}{1, 4}	3	
3	12	3	{1, 2}{2, 3}{3, 4}	3	
4	12	3	{1, 2}{2, 3}{4}	4	{X ₁ , X ₂ , X ₃ }{X ₄ }
4	3	2	{1, 2}{3, 4}	4	{X ₁ , X ₂ }{X ₃ , X ₄ }
5	6	3	{1, 2}{3}{4}	5	{X ₁ , X ₂ }{X ₃ }{X ₄ }
6	1	4	{1}{2}{3}{4}	6	{X ₁ }{X ₂ }{X ₃ }{X ₄ }

Dále uvedeme příklady grafů pro jednotlivé třídy modelů:

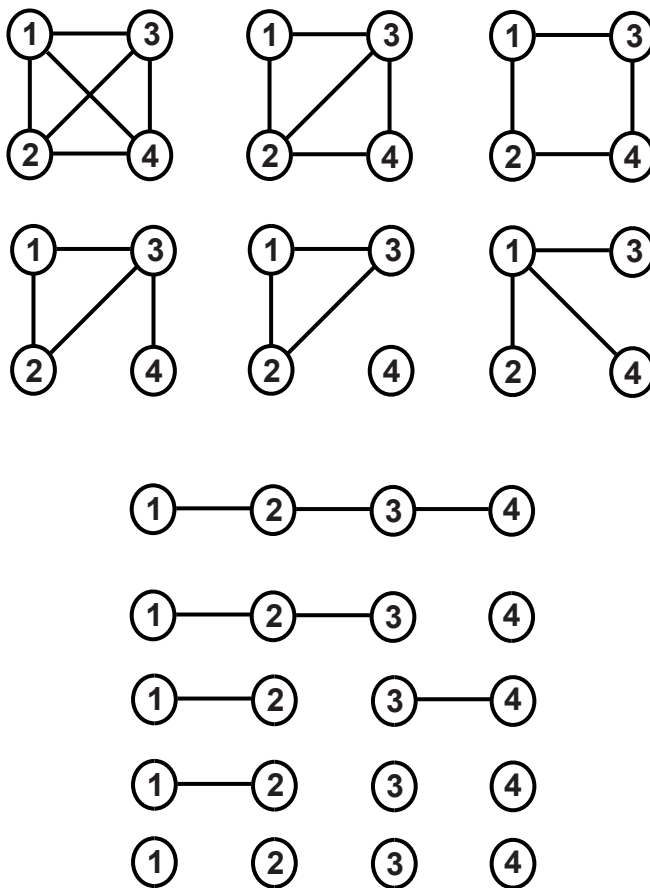
Dvousložkové náhodné vektory:



Třísložkové náhodné vektory:



Čtyřsložkové náhodné vektory:



Kapitola 2

Věrohodnostní přístup

Hlavním cílem této bakalářské práce je aplikovat grafické modely určené pro zpracování dat se spojitým rozdělením na data z finanční praxe. Pro tento účel jsme zvolili denní měnové kurzy z databáze ČNB [3] (na webové stránce www.cnb.cz). Pro výběr konkrétního grafického modelu byl použit program Backward1, vytvořený jako součást diplomové práce [4], realizovaný v systému Mathematica 4.0.

2.1 Zvolená metoda

Použitá data jsou ve formě realizace normálně rozděleného náhodného vektoru. Budeme se zabývat strukturou vzájemné závislosti jeho jednotlivých složek. Uvažujme všechny grafy podmíněných nezávislostí. Z nich vybereme ty, jež budou vykazovat nejlepší shodu s daty.

Hledáme tedy maximálně věrohodný odhad varianční matice \hat{V} , za podmínek daných grafickým modelem.

Výběr se jeví nejvýhodnější provést přímým výpočtem následujícím postupem, podrobně viz [5]:

- i. Určení systému k -rozměrných normálních rozdělení pro X a popis množiny parametrů. Pro mnohorozměrné normální rozdělení je nezávislost charakterizována varianční maticí $V = \text{var}(X)$ nebo její inverzí D .
- ii. Zvolíme grafický model, který budeme testovat na shodu s výše uvedenými daty.
- iii. Zkonstruujeme věrohodnostní funkci.
- iv. Určíme neznámé parametry. Maximalizujeme věrohodnostní funkci přes množinu neznámých parametrů. Požadavky na věrohodnostní funkci vyplývají z volby grafického modelu.

- v. Provedeme test shody vybraného grafického modelu s daty. Testová statistika (deviance) je rovna dvojnásobku rozdílu maximalizovaných logaritmických věrohodnostních funkcí. První je maximalizovaná bez omezení, druhá s omezením určeným vybraným grafickým modelem. Deviance má asymptoticky χ^2 -rozdělení, proto můžeme posoudit, zda grafický model datům vyhovuje.

IPF algoritmus

Problém velké citlivosti řešení na počáteční podmínku byl vyřešen v [7]. V této práci je popis IPF (Iterative Proportional Fitting) algoritmu. Jde o algoritmus iterační, který pro dvě hustoty g^0 a f k -rozměrného náhodného vektoru X hledá hustotu g^∞ , jež má stejnou interakční hustotu jako g^0 a identická marginální rozdělení jako f na podmnožinách množiny vrcholů $V = (v_1, \dots, v_k)$. Žádná z těchto podmnožin a_i nesmí být částí jiné, za podmínky

$$\bigcup_{i=1}^m a_i = V.$$

Toto splňují kliky grafu. Volíme je proto za příslušné podmnožiny. V n -tém kroku platí:

$$g_{ab}^{n+1} = g_{a|b}^n f_a.$$

Za a volíme v cyklu postupně kliky grafu. b určíme jako doplněk $b = V \setminus a$.

2.2 Testování vstupních dat

Při zpracování je nejprve nutné ověřit následující předpoklady požadované pro dané modely spojitých dat (provedeme při vlastním zpracování pomocí programu Backward1 [4]):

- i. Nezávislost logaritmických hodnot časové řady ($\log \frac{K_t}{K_{t-1}} = \log K_t - \log K_{t-1}$) pomocí testu založeného na znaménkách diferencí.
- ii. Normalita logaritmických kurzů

2.3 Selekční algoritmus

Pro výběr konkrétního grafického modelu s dobrou shodou s daty jsme použili program Backward1, součást diplomové práce [4], realizovaný v systému Mathematica 4.0. Program jsme upravili pro systém Mathematica 5.2.

Testovou statistikou je deviance. Vyhledávání odpovídajících grafů, které reprezentují konkrétní data, je popsáno v [7] na stranách 256–260. Jde o dva

tzv. Backward algoritmy vycházející z kompletního grafu. Postupně jsou z něj odebírány hrany dle určitého kritéria. V [7] jsou popsány také Forward algoritmy, které naopak vycházejí z grafu bez hran, a hrany do něj přidávají.

Backward1 – Backward algoritmus se stop pravidlem založeným na devianci vynechané hrany

Algoritmus je podrobně popsán v [4] na str. 31–34, programové řešení pak na str. 40–46. Backward algoritmus se stop pravidlem založeným na devianci vynechané hrany lze ve stručnosti popsat následujícími kroky:

- i. Výchozím je úplný graf daný saturovaným modelem.
- ii. Jeho deviance je položena rovna 0 ($dev^f = 0$).
- iii. Výpočet deviance vynechaných hran buď přímým výpočtem nebo pomocí IPF algoritmu.
- iv. Výběr nejmenší nevýznamné deviance a vyloučení příslušné hrany z grafu.
- v. Pokud jsou všechny deviance významné výpočet je ukončen.

Backward2 – Backward algoritmus se stop pravidlem založeným na celkové devianci

Tento algoritmus je podrobně popsán v [4] na str. 34–37. Algoritmus je identický s algoritmem z programu Backward1. Liší se jen ve STOP pravidle. Výpočet je totiž ukončen, pokud bude celková deviance modelu větší než příslušná kritická hodnota χ^2 -rozdělení.

Kapitola 3

Zpracování dat a interpretace výsledků

3.1 Postup přípravy dat ke zpracování

Výše uvedenou metodu jsme aplikovali na spojitá finanční data — denní kurzy devizového trhu ČNB [3]. Náhodný vektor, se kterým pracujeme, je v našem případě:

$$X = (1 \text{ CHF}, 1 \text{ EUR}, 1 \text{ GBP}, 100 \text{ JPY}, 1 \text{ USD}),$$

a to v řadě s denní frekvencí pro jednotlivé roky (1999–2005). Všechny kurzy se vztahují k české koruně.

Označení:

zkratka	měna
1 CHF	švýcarský frank
1 EUR	euro
1 GBP	britská libra
100 JPY	japonský jen
1 USD	americký dolar

- i. Ze souboru ročních měnových kurzů ČNB jsme vyseparovali vybrané měnové kurzy (CHF, EUR, GBP, JPY, USD).
- ii. Nahradili jsme časový údaj ve formě datumu odpovídající posloupností celých čísel značící pořadové číslo dne v roce.
- iii. Upravili jsme soubor importovaný z databáze ČNB, dle popisu z [4] str. 40, pomocí exportu z tabulkového procesoru MS Excel do textového souboru odděleného tabelátory.

- iv. Vstupní data jsou v textovém souboru, jednotlivé položky jsou odděleny tabelátory. V prvním řádku je uveden počet proměnných. Další řádek obsahuje popis položek, následují vlastní data. Dále uvedme vzorek zpracovávaných dat za rok 2005:

```

5
Date  1CHF  1EUR  1GBP  100JPY  1USD
3     19.663  30.365  42.936  21.873  22.484
4     19.634  30.385  43.026  21.941  22.736
5     19.601  30.430  43.171  22.039  23.009
6     19.643  30.440  43.247  22.013  23.090
7     19.567  30.305  43.118  21.950  22.965
10    19.558  30.250  43.351  22.111  23.089
11    19.586  30.300  43.269  22.164  23.053
.....
362   18.563  28.920  42.129  20.705  24.274
363   18.626  29.030  42.178  20.821  24.549
364   18.654  29.005  42.338  20.885  24.588

```

- v. Takto upravená data jsou připravena pro zpracování programem Backward1 pod systémem Mathematica 5.2.

3.2 Výsledky zpracování

Pro soubory měnových kurzů z let 1999–2005 jsme se pomocí programu Backward1 [4] pokoušeli nalézt konkrétní grafický model s dobrou shodou s daty. Protokoly ze zpracování pro jednotlivé roky jsou uvedeny v příloze č. 1. Rozmezí roků bylo zvoleno v souladu se vznikem evropské měny euro.

Jak je již výše zmíněno, byla zpracovávána řada pěti denních kurzů měn k české koruně.

Náhodný vektor byl tedy $X = (1 \text{ CHF}, 1 \text{ EUR}, 1 \text{ GBP}, 100 \text{ JPY}, 1 \text{ USD})$. Počet vrcholů grafu je 5. Počet všech možných grafů je 1024. Program Backward1 očísloval jednotlivé měny přirozenými čísly následujícím způsobem:

Označení	zkratka	měna
1	1 CHF	švýcarský frank
2	1 EUR	euro
3	1 GBP	britská libra
4	100 JPY	japonský jen
5	1 USD	americký dolar

Výsledky zpracování pro data z roku 2005

Varianční matice V zpracovávaných kurzů má tyto hodnoty:

$$\begin{pmatrix} 0.0000136516 & 0.0000106439 & 0.0000104458 & 0.000011017 & 0.0000107111 \\ 0.0000106439 & 0.0000104219 & 0.0000092897 & 0.000009408 & 0.0000116232 \\ 0.0000104458 & 0.0000092897 & 0.0000191808 & 0.000015053 & 0.0000190967 \\ 0.0000110170 & 0.0000094082 & 0.0000150536 & 0.000029023 & 0.0000227082 \\ 0.0000107111 & 0.0000116232 & 0.0000190967 & 0.000022708 & 0.0000427649 \end{pmatrix}$$

Korelační matice je následující:

$$\begin{pmatrix} 1 & 0.892355 & 0.645532 & 0.553418 & 0.443303 \\ 0.892355 & 1 & 0.657051 & 0.540897 & 0.550566 \\ 0.645532 & 0.657051 & 1 & 0.637954 & 0.666777 \\ 0.553418 & 0.540897 & 0.637954 & 1 & 0.644497 \\ 0.443303 & 0.550566 & 0.666777 & 0.644497 & 1 \end{pmatrix}$$

Je setavena tato matice:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Znaménko „-“ značí vynechanou hranu.

Číslo k u znaménka znamená, že hrana byla vynechána v k -tém kroku.

Jsou vynechány 2 hrany mezi vrcholy reprezentujícími: 2(EUR)-3(GBP), 2(EUR)-4(JPY).

Z grafu je tedy patrná podmíněná nezávislost ve dvojicích měn.

Celková deviance je 3.2926.

p -hodnota testu je 0.192762.

Nezamítáme tedy shodu modelu s daty.

Pro data z předcházejících let 1999–2004, která byla zpracovávána, uvedme následující parametry:

i. Rok 1999:

Korelační matice :

$$\begin{pmatrix} 1 & 0.938629 & 0.651140 & 0.364175 & 0.492329 \\ 0.938629 & 1 & 0.706102 & 0.379420 & 0.561360 \\ 0.651140 & 0.706102 & 1 & 0.493834 & 0.773660 \\ 0.364175 & 0.379420 & 0.493834 & 1 & 0.546675 \\ 0.492329 & 0.561360 & 0.773660 & 0.546675 & 1 \end{pmatrix}$$

Je vynecháno 5 hran: 1(CHF)–3(GBP), 1(CHF)–4(JPY), 1(CHF)–5(USD), 2(EUR)–4(JPY), 2(EUR)–5(USD)

Celková deviance je 6.69108.

p -hodnota testu je 0.244648.

ii. Rok 2000:

Korelační matice :

$$\begin{pmatrix} 1 & 0.7012670 & 0.348899 & 0.2511430 & 0.3067050 \\ 0.701267 & 1 & 0.211062 & 0.0785891 & 0.0818733 \\ 0.348899 & 0.2110620 & 1 & 0.5349110 & 0.7097160 \\ 0.251143 & 0.0785891 & 0.534911 & 1 & 0.7759210 \\ 0.306705 & 0.0818733 & 0.709716 & 0.7759210 & 1 \end{pmatrix}$$

Jsou vynechány 4 hrany: 1(CHF)–3(GBP), 1(CHF)–4(JPY), 2(EUR)–4(JPY), 3(GBP)–4(JPY)

Celková deviance je 1.47457.

p -hodnota testu je 0.831137.

iii. Rok 2001:

Korelační matice :

$$\begin{pmatrix} 1 & 0.794988 & 0.388516 & 0.371527 & 0.300257 \\ 0.794988 & 1 & 0.416757 & 0.224353 & 0.248770 \\ 0.388516 & 0.416757 & 1 & 0.530616 & 0.721726 \\ 0.371527 & 0.224353 & 0.530616 & 1 & 0.680527 \\ 0.300257 & 0.248770 & 0.721726 & 0.680527 & 1 \end{pmatrix}$$

Jsou vynechány 4 hrany: 1(CHF)–3(GBP), 1(CHF)–5(USD), 2(EUR)–5(USD), 3(GBP)–4(JPY)

Celková deviance je 5.20233.

p -hodnota testu je 0.26716.

iv. **Rok 2002:**

Korelační matice :

$$\begin{pmatrix} 1 & 0.943868 & 0.756004 & 0.665344 & 0.488962 \\ 0.943868 & 1 & 0.784594 & 0.692879 & 0.535397 \\ 0.756004 & 0.784594 & 1 & 0.676487 & 0.780013 \\ 0.665344 & 0.692879 & 0.676487 & 1 & 0.601253 \\ 0.488962 & 0.535397 & 0.780013 & 0.601253 & 1 \end{pmatrix}$$

Jsou vynechány 2 hrany: 2(EUR)–5(USD), 3(GBP)–4(JPY)

Celková deviance je 2.94739.

p -hodnota testu je 0.229078.

v. **Rok 2003:**

Korelační matice :

$$\begin{pmatrix} 1 & 0.815475 & 0.411939 & 0.385859 & 0.298527 \\ 0.815475 & 1 & 0.516307 & 0.452973 & 0.395921 \\ 0.411939 & 0.516307 & 1 & 0.538691 & 0.702946 \\ 0.385859 & 0.452973 & 0.538691 & 1 & 0.714007 \\ 0.298527 & 0.395921 & 0.702946 & 0.714007 & 1 \end{pmatrix}$$

Je vynechána 1 hrana: 2(EUR)–5(USD)

Celková deviance je 0.00253054.

p -hodnota testu je 0.95988.

vi. **Rok 2004:**

Korelační matice :

$$\begin{pmatrix} 1 & 0.818333 & 0.488971 & 0.347380 & 0.386622 \\ 0.818333 & 1 & 0.634790 & 0.404467 & 0.502386 \\ 0.488971 & 0.634790 & 1 & 0.433724 & 0.480440 \\ 0.347380 & 0.404467 & 0.433724 & 1 & 0.622169 \\ 0.386622 & 0.502386 & 0.480440 & 0.622168 & 1 \end{pmatrix}$$

Jsou vynechány 4 hrany: 1(CHF)–3(GBP), 1(CHF)–4(JPY), 1(CHF)–5(USD), 2(EUR)–4(JPY)

Celková deviance je 3.41686.

p -hodnota testu je 0.490633.

Grafy pro jednotlivé roky jsou nakresleny v příloze.

Ve všech grafech za jednotlivé roky se vyskytují hrany mezi vrcholy 1(CHF)–2(EUR), 3(GBP)–5(USD) a 4(JPY)–5(USD), což odpovídá podmíněné závislosti kurzů těchto měn. Ve všech ročních zpracováních kromě roku 2005 se též vyskytuje hrana mezi vrcholy 2(EUR)–3(GBP). Naopak každá z hran se v některém z grafů vyskytuje alespoň jednou. Meziročně se grafy poměrně výrazně mění, žádné dva nejsou identické.

Možnou ekonomickou interpretací těchto změn je zánik evropských měn, které nahradilo euro, s nimiž až do roku 2002 existovalo paralelně (BEF, FIM, FRF, IEP, ITL, LUF, NLG, PTE, ATS, DEM, ESP). Na podmíněné závislosti kurzu USD k ostatním měnám mohlo mít vliv i rozkolísání kurzu USD v důsledku války v Iráku v roce 2004.

V některých grafech lze také studovat efekty separace. Zmiňme rok 1999, kde 3(GBP) separuje {1(CHF), 2(EUR)} a {4(JPY), 5(USD)}. Kurzy měn z kontinentální Evropy při vyloučení vlivu britské libry jsou tedy podmíněně nezávislé s kurzy mimoevropských měn.

Kapitola 4

Závěr

4.1 Interpretace výsledků

Hlavním cílem práce byla aplikace grafických modelů na spojitá data z finanční praxe — kurzy pěti měn v ročních souborech za období pěti let. Pro selekci modelu byl použit program Backward1 [4] upravený pro verzi programu Mathematica 5.2. Pomocí algoritmu tohoto programu byly vyhledány grafické modely s grafem dobře reprezentujícím vstupní data.

Protokoly ze zpracování jsou uvedeny v příloze č. 1. Byly nalezeny podmíněné závislosti, které se vyskytují ve všech ročních zpracováních. Každá hrana se za sledované období 1999–2005 vyskytla v grafech alespoň jednou. Zmínili jsme i možné příčiny značné diverzity mezi ročními grafy.

V žádném z grafů se neobjevují nepodmíněně nezávislé složky ani podvektory vektorů měnových kurzů. Lze tedy konstatovat provázanost jednotlivých kurzů ve sledovaném časovém horizontu.

4.2 Poznámky k výpočtu denních kurzů ČNB

Uveďme, jakým způsobem se počítají průměrné devizové kurzy ČNB [3]. Kromě denních kurzů devizového trhu publikuje Česká národní banka také průměrné devizové kurzy za jednotlivé měsíce, čtvrtletí, kumulovaná čtvrtletí a roky. Zveřejněná hodnota je pak aritmetickým průměrem z denních kurzů vyhlášených v pracovní dny za dané období. Průměrné kurzy za uvedená období se zveřejňují současně se zveřejněním posledního denního kurzu spadajícího do příslušného období. Matematické vyjádření je:

$$Y = \frac{1}{n(i)} \sum_i x_i$$

- Y výsledný průměrný kurz za dané období;
- i kurz vyhlášený k pracovnímu dni i ;
- $n(i)$ počet pracovních dnů i v daném období.

Nejsou-li k dispozici všechny denní kurzy příslušné měny za dané období (tj. pokud kurz nebyl v nějakém období uváděn v kurzovním lístku ČNB), průměrný kurz se nevypočte a do příslušného pole se neuvede žádná hodnota.

Literatura

- [1] Anděl, J.: *Základy matematické statistiky*, Matfyzpress, Praha, 2005.
- [2] Cipra, T.: *Analýza časových řad s aplikacemi v ekonomii*, SNTL, Praha, 1986.
- [3] Kurzovní lístek ČNB: <http://www.cnb.cz> .
- [4] Chýna, V.: *Grafické modely pro analýzu spojitých finančních dat*, Diplomová práce, MFF UK, Praha, 2002.
- [5] Lněnička, R.: *Bayesovské metody pro zpracování finančních dat*, Diplomová práce, MFF UK, Praha, 2001.
- [6] Matoušek, J., Nešetřil, J.: *Kapitoly z diskrétní matematiky*, Karolinum, Praha, 2002.
- [7] Whittaker, J.: *Graphical Models in Applied Multivariate Statistics*, Willey, New York, 1990.

Příloha:

Protokoly zpracování programem Backward1 (1999–2005)

Year 1999

```
1CHF      0.0651964
1EUR      0.328972
p-value testu znamenek diferenci: 1GBP      0.0651964
100JPY    0.0651964
1USD      0.587594
```

Zvolene indexy:

```
1 1CHF
2 1EUR
3 1GBP
4 100JPY
5 1USD
```

Pocet realizaci = delka log vynosu =254

pocet vrcholu = 5

```
variancni matice=
      0.0000254357 0.0000212379 0.0000209684 0.00001896 0.0000177147
      0.0000212379 0.0000201277 0.0000202271 0.0000175721 0.0000179678
      0.0000209684 0.0000202271 0.0000407698 0.0000325504 0.0000352433
      0.00001896 0.0000175721 0.0000325504 0.000106565 0.0000402617
      0.0000177147 0.0000179678 0.0000352433 0.0000402617 0.0000508995
```

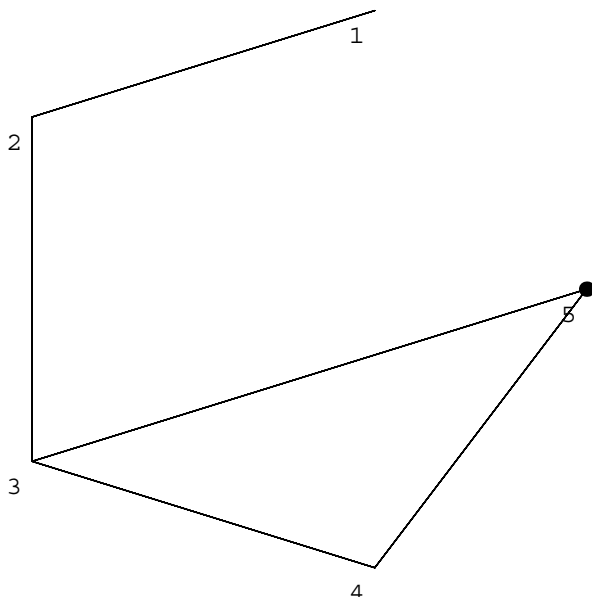
```
korel matice=
      1. 0.938629 0.65114 0.364175 0.492329
      0.938629 1. 0.706102 0.37942 0.56136
      0.65114 0.706102 1. 0.493834 0.77366
      0.364175 0.37942 0.493834 1. 0.546675
      0.492329 0.56136 0.77366 0.546675 1.
```

```
matice souslednosti =
      0 0 0 0 0
      1 0 0 0 0
      -1 1 0 0 0
      -3 -2 1 0 0
      -4 -5 1 1 0
```

Pocet vynechanych hran = 5

Celkova deviance = 6.69108

p-value = 0.244648



Year 2000

	1CHF	0.742928
	1EUR	0.325148
p-value testu znamenek diferenci:	1GBP	0.229136
	100JPY	0.101028
	1USD	0.584633

Zvolene indexy:

1	1CHF
2	1EUR
3	1GBP
4	100JPY
5	1USD

Pocet realizaci = delka log vynosu =250

pocet vrcholu = 5

variancni matice=	0.0000131194	$7.06701 \cdot 10^{-6}$	$8.87511 \cdot 10^{-6}$	$9.08039 \cdot 10^{-6}$	$8.64992 \cdot 10^{-6}$
	$7.06701 \cdot 10^{-6}$	$7.7409 \cdot 10^{-6}$	$4.12405 \cdot 10^{-6}$	$2.18265 \cdot 10^{-6}$	$1.77367 \cdot 10^{-6}$
	$8.87511 \cdot 10^{-6}$	$4.12405 \cdot 10^{-6}$	0.0000493213	0.0000374995	0.0000388093
	$9.08039 \cdot 10^{-6}$	$2.18265 \cdot 10^{-6}$	0.0000374995	0.0000996443	0.0000603084
	$8.64992 \cdot 10^{-6}$	$1.77367 \cdot 10^{-6}$	0.0000388093	0.0000603084	0.0000606272

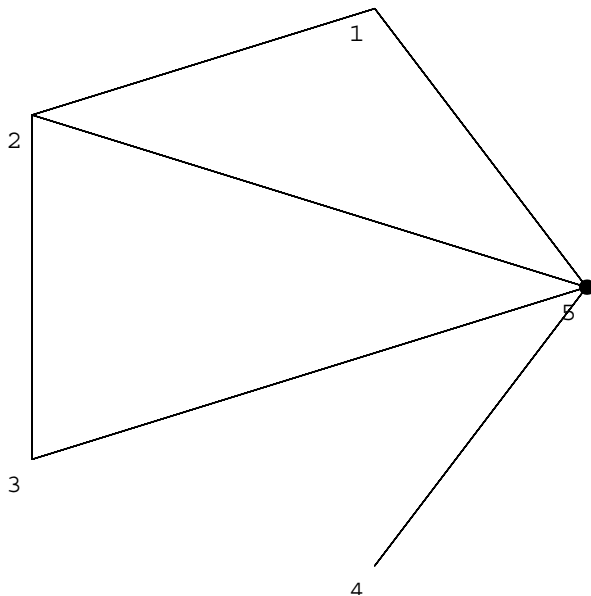
korel matice=	1.	0.701267	0.348899	0.251143	0.306705
	0.701267	1.	0.211062	0.0785891	0.0818733
	0.348899	0.211062	1.	0.534911	0.709716
	0.251143	0.0785891	0.534911	1.	0.775921
	0.306705	0.0818733	0.709716	0.775921	1.

matice souslednosti =	0	0	0	0	0
	1	0	0	0	0
	-4	1	0	0	0
	-1	-2	-3	0	0
	1	1	1	1	0

Pocet vynechanych hran = 4

Celkova deviance = 1.47457

p-value = 0.831137



Year 2001

	1CHF	0.444103
	1EUR	0.444103
p-value testu znamenek diferenci:	1GBP	0.742928
	100JPY	0.101028
	1USD	0.155247

Zvolene indexy:

1	1CHF
2	1EUR
3	1GBP
4	100JPY
5	1USD

Pocet realizaci = delka log vynosu =250

pocet vrcholu = 5

variancni matice=	0.0000208201	0.0000123274	$9.88431 \cdot 10^{-6}$	0.0000149537	$9.61294 \cdot 10^{-6}$
	0.0000123274	0.0000115488	$7.89673 \cdot 10^{-6}$	$6.72538 \cdot 10^{-6}$	$5.9318 \cdot 10^{-6}$
	$9.88431 \cdot 10^{-6}$	$7.89673 \cdot 10^{-6}$	0.0000310879	0.0000260971	0.0000282351
	0.0000149537	$6.72538 \cdot 10^{-6}$	0.0000260971	0.0000778094	0.0000421195
	$9.61294 \cdot 10^{-6}$	$5.9318 \cdot 10^{-6}$	0.0000282351	0.0000421195	0.0000492315

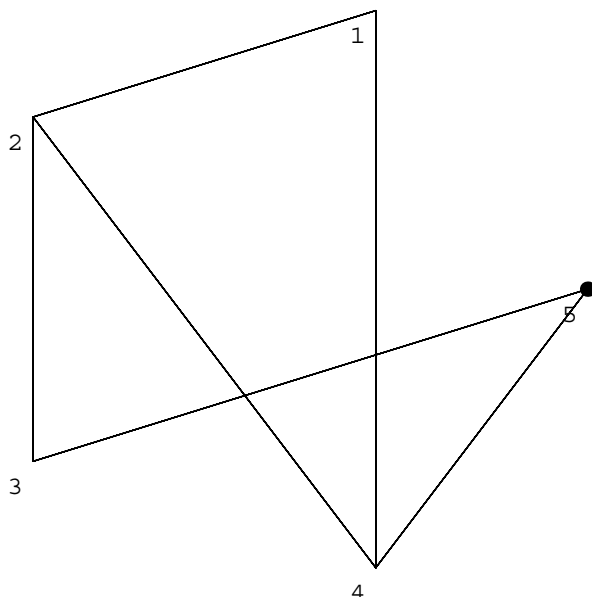
korel matice=	1.	0.794988	0.388516	0.371527	0.300257
	0.794988	1.	0.416757	0.224353	0.24877
	0.388516	0.416757	1.	0.530616	0.721726
	0.371527	0.224353	0.530616	1.	0.680527
	0.300257	0.24877	0.721726	0.680527	1.

matice souslednosti =	0	0	0	0	0
	1	0	0	0	0
	-2	1	0	0	0
	1	1	-3	0	0
	-1	-4	1	1	0

Pocet vynechanych hran = 4

Celkova deviance = 5.20233

p-value = 0.26716



Year 2002

	1CHF	0.382733
	1EUR	0.275234
p-value testu znamenek diferenci:	1GBP	0.512691
	100JPY	0.662521
	1USD	0.0290963

Zvolene indexy:

1	1CHF
2	1EUR
3	1GBP
4	100JPY
5	1USD

Pocet realizaci = delka log vynosu =251

pocet vrcholu = 5

	0.0000303877	0.0000251057	0.0000217019	0.0000264682	0.0000182239
	0.0000251057	0.0000232822	0.0000197143	0.0000241268	0.0000174665
variancni matice=	0.0000217019	0.0000197143	0.0000271175	0.0000254222	0.0000274628
	0.0000264682	0.0000241268	0.0000254222	0.0000520786	0.0000293363
	0.0000182239	0.0000174665	0.0000274628	0.0000293363	0.0000457127

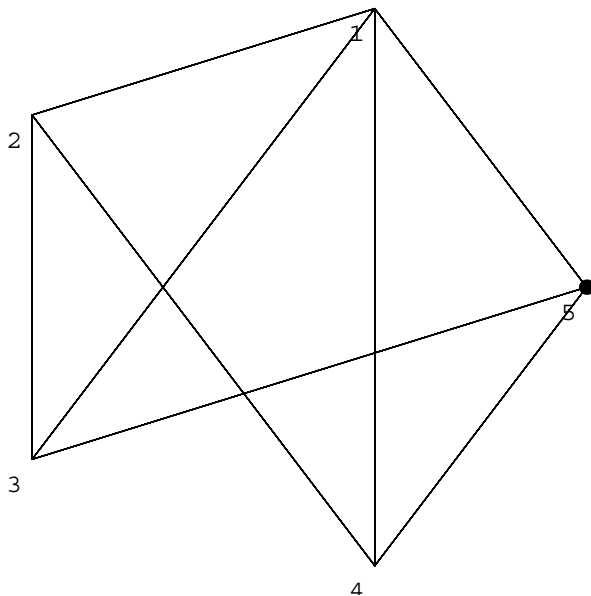
	1.	0.943868	0.756004	0.665344	0.488962
	0.943868	1.	0.784594	0.692879	0.535397
korel matice=	0.756004	0.784594	1.	0.676487	0.780013
	0.665344	0.692879	0.676487	1.	0.601253
	0.488962	0.535397	0.780013	0.601253	1.

	0	0	0	0	0
	1	0	0	0	0
matice souslednosti =	1	1	0	0	0
	1	1	-2	0	0
	1	-1	1	1	0

Pocet vynechanych hran = 2

Celkova deviance = 2.94739

p-value = 0.229078



Year 2003

	1CHF	0.000480341
	1EUR	0.12663
p-value testu znamenek diferenci:	1GBP	0.512691
	100JPY	1.
	1USD	0.275234

Zvolene indexy:

1	1CHF
2	1EUR
3	1GBP
4	100JPY
5	1USD

Pocet realizaci = delka log vynosu =251

pocet vrcholu = 5

variancni matice=	0.000017682	0.0000116304	$8.95093 \cdot 10^{-6}$	0.0000106541	$8.86751 \cdot 10^{-6}$
	0.0000116304	0.0000115036	$9.04885 \cdot 10^{-6}$	0.0000100882	$9.48587 \cdot 10^{-6}$
	$8.95093 \cdot 10^{-6}$	$9.04885 \cdot 10^{-6}$	0.0000267016	0.0000182782	0.0000256592
	0.0000106541	0.0000100882	0.0000182782	0.000043117	0.0000331191
	$8.86751 \cdot 10^{-6}$	$9.48587 \cdot 10^{-6}$	0.0000256592	0.0000331191	0.0000499004

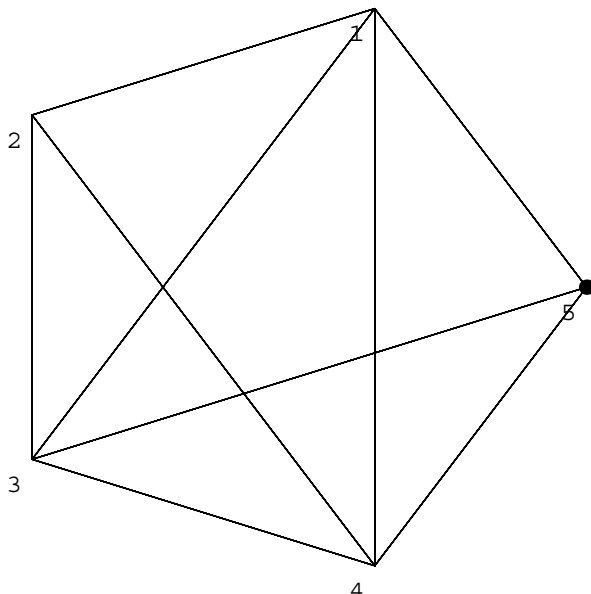
korel matice=	1.	0.815475	0.411939	0.385859	0.298527
	0.815475	1.	0.516307	0.452973	0.395921
	0.411939	0.516307	1.	0.538691	0.702946
	0.385859	0.452973	0.538691	1.	0.714007
	0.298527	0.395921	0.702946	0.714007	1.

matice souslednosti =	0	0	0	0	0
	1	0	0	0	0
	1	1	0	0	0
	1	1	1	0	0
	1	-1	1	1	0

Pocet vynechanych hran = 1

Celkova deviance = 0.00253054

p-value = 0.95988



Year 2004

	1CHF	0.08206
	1EUR	0.384613
p-value testu znamenek diferenci:	1GBP	0.514355
	100JPY	0.66377
	1USD	1.

Zvolene indexy:

1	1CHF
2	1EUR
3	1GBP
4	100JPY
5	1USD

Pocet realizaci = delka log vynosu =253

pocet vrcholu = 5

variancni matice=	0.0000133507	$9.39509 \cdot 10^{-6}$	$8.93726 \cdot 10^{-6}$	$8.61932 \cdot 10^{-6}$	$9.84137 \cdot 10^{-6}$
	$9.39509 \cdot 10^{-6}$	$9.87275 \cdot 10^{-6}$	$9.97742 \cdot 10^{-6}$	$8.63015 \cdot 10^{-6}$	0.000010997
	$8.93726 \cdot 10^{-6}$	$9.97742 \cdot 10^{-6}$	0.000025023	0.0000147333	0.0000167427
	$8.61932 \cdot 10^{-6}$	$8.63015 \cdot 10^{-6}$	0.0000147333	0.0000461139	0.0000294335
	$9.84137 \cdot 10^{-6}$	0.000010997	0.0000167427	0.0000294335	0.0000485329

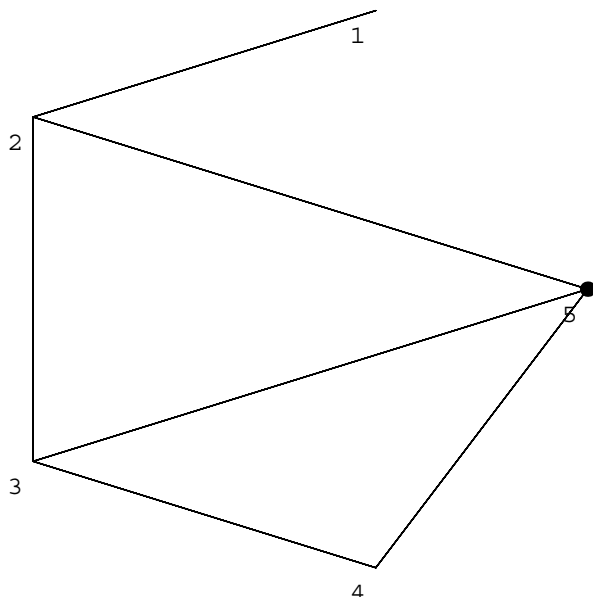
korel matice=	1.	0.818333	0.488971	0.34738	0.386622
	0.818333	1.	0.63479	0.404467	0.502386
	0.488971	0.63479	1.	0.433724	0.48044
	0.34738	0.404467	0.433724	1.	0.622168
	0.386622	0.502386	0.48044	0.622168	1.

matice souslednosti =	0	0	0	0	0
	1	0	0	0	0
	-4	1	0	0	0
	-2	-1	1	0	0
	-3	1	1	1	0

Pocet vynechanych hran = 4

Celkova deviance = 3.41686

p-value = 0.490633



Year 2005

	1CHF	0.327067
	1EUR	0.512691
p-value testu znamenek diferenci:	1GBP	0.586121
	100JPY	0.230985
	1USD	0.230985

Zvolene indexy:

1	1CHF
2	1EUR
3	1GBP
4	100JPY
5	1USD

Pocet realizaci = delka log vynosu =252

pocet vrcholu = 5

```
variancni matice=
      0.0000136516  0.0000106439  0.0000104458  0.000011017  0.0000107111
      0.0000106439  0.0000104219  9.28978 · 10-6  9.40819 · 10-6  0.0000116232
      0.0000104458  9.28978 · 10-6  0.0000191808  0.0000150536  0.0000190967
      0.000011017  9.40819 · 10-6  0.0000150536  0.0000290293  0.0000227082
      0.0000107111  0.0000116232  0.0000190967  0.0000227082  0.0000427649
```

```
korel matice=
      1.  0.892355  0.645532  0.553418  0.443303
      0.892355  1.  0.657051  0.540897  0.550566
      0.645532  0.657051  1.  0.637954  0.666777
      0.553418  0.540897  0.637954  1.  0.644497
      0.443303  0.550566  0.666777  0.644497  1.
```

```
matice souslednosti =
      0  0  0  0  0
      1  0  0  0  0
      1 -1  0  0  0
      1 -2  1  0  0
      1  1  1  1  0
```

Pocet vynechanych hran = 2

Celkova deviance = 3.2926

p-value = 0.192762

