

Estimating correlation from OHLC prices

Jozef Kiseľák – Štefan Lyócsa – Tomáš Výrost

Analytické metódy a Big Data manažment v praxi
7.12.2015

Úloha

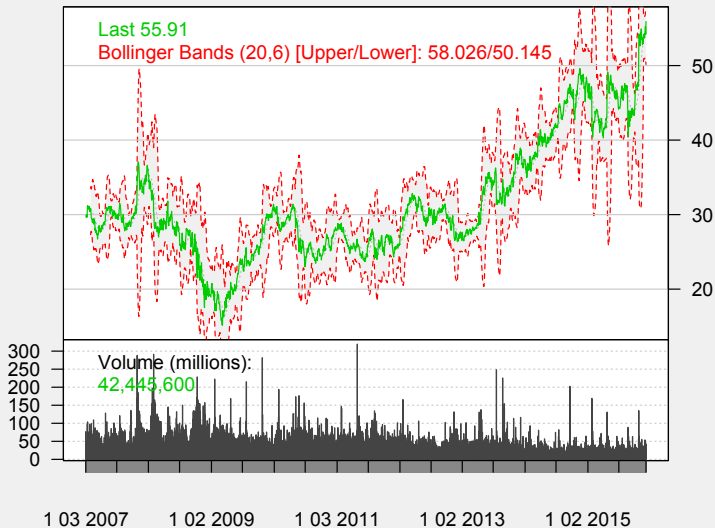
- Skúmanie vývoja cien **finančných investícií**.
- Z hľadiska tvorby diverzifikovaného portfólia nás zaujímajú **vzťahy** medzi výnosmi investícií.

“Lepšie informácie o cenách umožnia lepšie modely”

- **Zvyšovanie frekvencie** zbieraných údajov – problémy:
 - komplexita – Big Data
 - čo ak dáta nie sú na požadovanej škále dostupné?
- Riešenie problémov lepšími **analytickými metódami**
 - ako z dostupných údajov získať viac?

MSFT

[2007-01-03/2015-12-04]

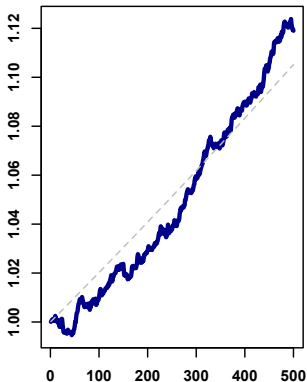


Dvozmerný Brownov pohyb s driftom

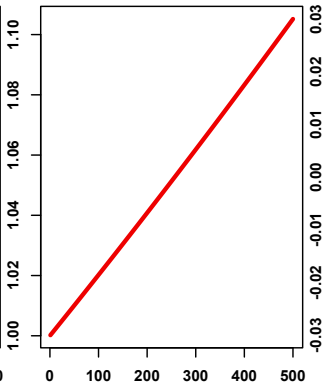
Skúmame dva procesy X_t^1 a X_t^2 , ktoré sa riadia geometrickým Brownovým pohybom. Formálne pre $j = 1, 2$:

$$\frac{dX_t^j}{X_t^j} = \tilde{\mu}_j dt + \sigma_j dW_t^j$$

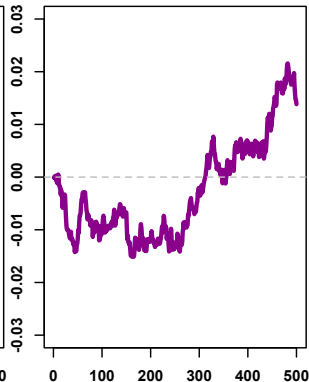
Celý proces



Deterministická časť



Stochastická časť



Dvojrozmerný Brownov pohyb s driftom

Skúmame dva procesy X_t^1 a X_t^2 , ktoré sa riadia geometrickým Brownovým pohybom. Formálne pre $j = 1, 2$:

$$\frac{dX_t^j}{X_t^j} = \tilde{\mu}_j dt + \sigma_j dW_t^j$$

Ďalej predpokladajme, že štandardné Wienerove procesy W_t^1 a W_t^2 sú korelované: $\mathbb{E}[W_t^1 W_s^2] = \rho \min(s, t)$

Je známe, že pre takto definovaný proces

$$\mathbb{E}[X_t^j] = X_0^j e^{\tilde{\mu}_j t}, \quad \text{Var}(X_t^j) = (X_0^j)^2 e^{2\tilde{\mu}_j t} (e^{\sigma_j^2 t} - 1)$$

Následne pre vzájomnú koreláciu dostávame

$$\rho_{X_t^1, X_t^2} := \text{cor}(X_t^1, X_t^2) = \frac{e^{\rho \sigma_1 \sigma_2 t} - 1}{\sqrt{e^{\sigma_1^2 t} - 1} \sqrt{e^{\sigma_2^2 t} - 1}} \xrightarrow{t \rightarrow 0^+} \rho, \quad \forall \sigma_1, \sigma_2 > 0.$$

Dvojrozmerný Brownov pohyb s driftom

Skúmame dva procesy X_t^1 a X_t^2 , ktoré sa riadia geometrickým Brownovým pohybom. Formálne pre $j = 1, 2$:

$$\frac{dX_t^j}{X_t^j} = \tilde{\mu}_j dt + \sigma_j dW_t^j$$

Ďalej predpokladajme, že štandardné Wienerove procesy W_t^1 a W_t^2 sú korelované: $\mathbb{E}[W_t^1 W_s^2] = \rho \min(s, t)$

Je známe, že pre takto definovaný proces

$$\mathbb{E}[X_t^j] = X_0^j e^{\tilde{\mu}_j t}, \quad \text{Var}(X_t^j) = (X_0^j)^2 e^{2\tilde{\mu}_j t} (e^{\sigma_j^2 t} - 1)$$

Následne pre vzájomnú koreláciu dostávame

$$\rho_{X_t^1, X_t^2} := \text{cor}(X_t^1, X_t^2) = \frac{e^{\rho \sigma_1 \sigma_2 t} - 1}{\sqrt{e^{\sigma_1^2 t} - 1} \sqrt{e^{\sigma_2^2 t} - 1}} \xrightarrow{t \rightarrow 0^+} \rho, \quad \forall \sigma_1, \sigma_2 > 0.$$

Dvojrozmerný Brownov pohyb s driftom

Skúmame dva procesy X_t^1 a X_t^2 , ktoré sa riadia geometrickým Brownovým pohybom. Formálne pre $j = 1, 2$:

$$\frac{dX_t^j}{X_t^j} = \tilde{\mu}_j dt + \sigma_j dW_t^j$$

Ďalej predpokladajme, že štandardné Wienerove procesy W_t^1 a W_t^2 sú korelované: $\mathbb{E}[W_t^1 W_s^2] = \rho \min(s, t)$

Je známe, že pre takto definovaný proces

$$\mathbb{E}[X_t^j] = X_0^j e^{\tilde{\mu}_j t}, \quad \text{Var}(X_t^j) = (X_0^j)^2 e^{2\tilde{\mu}_j t} (e^{\sigma_j^2 t} - 1)$$

Následne pre vzájomnú koreláciu dostávame

$$\rho_{X_t^1, X_t^2} := \text{cor}(X_t^1, X_t^2) = \frac{e^{\rho \sigma_1 \sigma_2 t} - 1}{\sqrt{e^{\sigma_1^2 t} - 1} \sqrt{e^{\sigma_2^2 t} - 1}} \xrightarrow{t \rightarrow 0^+} \rho, \quad \forall \sigma_1, \sigma_2 > 0.$$

Skúmame dva procesy X_t^1 a X_t^2 , ktoré sa riadia geometrickým Brownovým pohybom. Formálne pre $j = 1, 2$:

$$\frac{dX_t^j}{X_t^j} = \tilde{\mu}_j dt + \sigma_j dW_t^j$$

Ďalej predpokladajme, že štandardné Wienerove procesy W_t^1 a W_t^2 sú korelované: $\mathbb{E}[W_t^1 W_s^2] = \rho \min(s, t)$

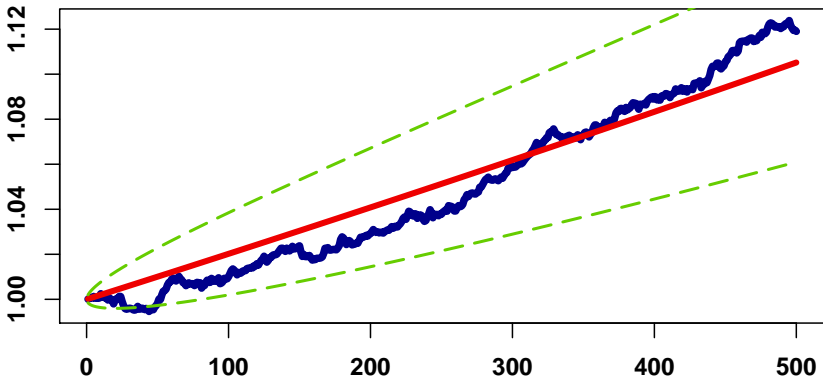
Je známe, že pre takto definovaný proces

$$\mathbb{E}[X_t^j] = X_0^j e^{\tilde{\mu}_j t}, \quad \text{Var}(X_t^j) = (X_0^j)^2 e^{2\tilde{\mu}_j t} (e^{\sigma_j^2 t} - 1)$$

Následne pre vzájomnú koreláciu dostávame

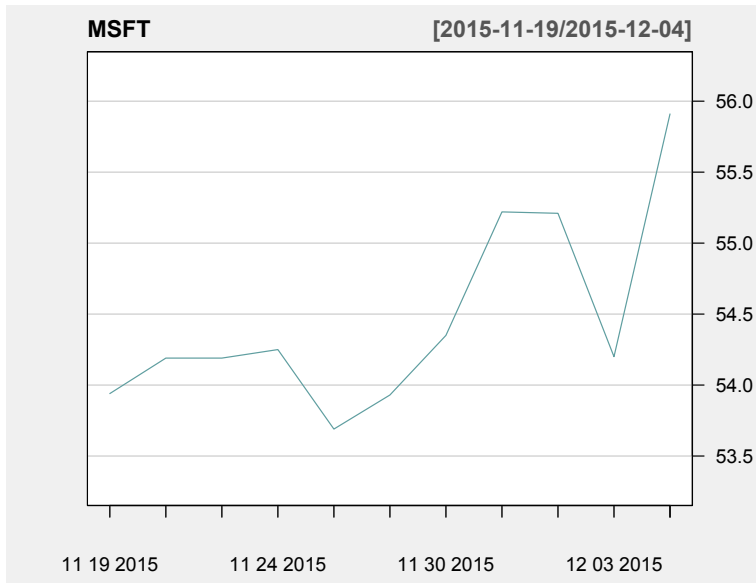
$$\rho_{X_t^1, X_t^2} := \text{cor}(X_t^1, X_t^2) = \frac{e^{\rho \sigma_1 \sigma_2 t} - 1}{\sqrt{e^{\sigma_1^2 t} - 1} \sqrt{e^{\sigma_2^2 t} - 1}} \xrightarrow{t \rightarrow 0^+} \rho, \quad \forall \sigma_1, \sigma_2 > 0.$$

Dvojrozmerný Brownov pohyb s driftom



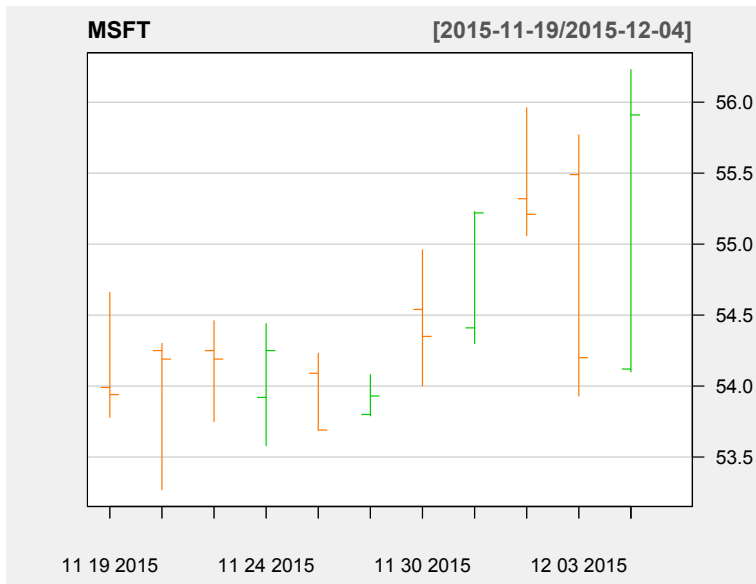
Obchodovanie v definovaných časových intervaloch

Line chart



Obchodovanie v definovaných časových intervaloch

Bar chart



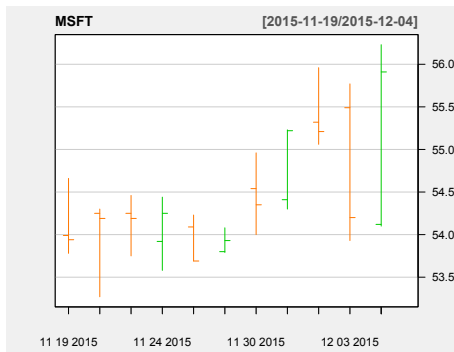
Diskretizácia času a prechod k OHLC

Uvažujme o obchodovaní v časových intervaloch

$t_1 < t_2 < \dots < t_n$ o dĺžke δ_i , kde $t_i < t_i + \delta_i$, $i = 1, \dots, n - 1$.

Pre takto diskretizovaný proces môžeme definovať otváraciu, uzatváraciu, najvyššiu a najnižšiu cenu pre periódu i :

$$O_i^j = X_{t_i}, \quad C_i^j = X_{t_i + \delta_i}, \quad H_i^j = \max_{t_i \leq t \leq t_i + \delta_i} X_t^j, \quad L_i^j = \min_{t_i \leq t \leq t_i + \delta_i} X_t^j.$$



Spojité výnosnosti v rámci časového intervalu i :

$$R_i^j = \ln \frac{C_i^j}{O_i^j} \sim N(\mu_j \delta_i, \sigma_j^2 \delta_i)$$

Zaujímavé je však to, či niečo dokážeme povedať o vzťahu medzi výnosnosťami daných dvoch aktív:

$$\hat{\rho}_C = \widehat{\text{cor}}(R_i^1, R_i^2) \xrightarrow{D} N\left(\rho, \frac{(1 - \rho^2)^2}{n}\right)$$

Nevýhody: pomalá konvergencia pre $|\rho| \approx 1$ a závislosť rozptylu na ρ . Používa sa preto Fisherova z-transformácia

$$\frac{1}{2} \ln \left(\frac{1 + \hat{\rho}_C}{1 - \hat{\rho}_C} \right)$$

Spojité výnosnosť v rámci časového intervalu i :

$$R_i^j = \ln \frac{C_i^j}{O_i^j} \sim N(\mu_j \delta_i, \sigma_j^2 \delta_i)$$

Zaujímavé je však to, či niečo dokážeme povedať o vzťahu medzi výnosnosťami daných dvoch aktív:

$$\hat{\rho}_C = \widehat{\text{cor}}(R_i^1, R_i^2) \xrightarrow{D} N\left(\rho, \frac{(1 - \rho^2)^2}{n}\right)$$

Nevýhody: pomalá konvergencia pre $|\rho| \approx 1$ a závislosť rozptylu na ρ . Používa sa preto Fisherova z-transformácia

$$\frac{1}{2} \ln \left(\frac{1 + \hat{\rho}_C}{1 - \hat{\rho}_C} \right)$$

Spojité výnosnosť v rámci časového intervalu i :

$$R_i^j = \ln \frac{C_i^j}{O_i^j} \sim N(\mu_j \delta_i, \sigma_j^2 \delta_i)$$

Zaujímavé je však to, či niečo dokážeme povedať o vzťahu medzi výnosnosťami daných dvoch aktív:

$$\hat{\rho}_C = \widehat{\text{cor}}(R_i^1, R_i^2) \xrightarrow{D} N\left(\rho, \frac{(1 - \rho^2)^2}{n}\right)$$

Nevýhody: pomalá konvergencia pre $|\rho| \approx 1$ a závislosť rozptylu na ρ . Používa sa preto Fisherova z-transformácia

$$\frac{1}{2} \ln \left(\frac{1 + \hat{\rho}_C}{1 - \hat{\rho}_C} \right)$$

Podľa Itoovej lemy sa logaritmy cien $Y_t^j = \ln(X_t^j)$ vyvíjajú ako Brownov pohyb

$$dY_t^j = \mu_j dt + \sigma_j dW_t^j, \quad j = 1, 2.$$

Definujme procesy

$$Z_t^j = Y_t^j - L_t^j = Y_t^j - \inf_{0 \leq s \leq t} Y_s^j \quad \text{a} \quad z_t^j = Y_t^j - H_t^j = Y_t^j - \sup_{0 \leq s \leq t} Y_s^j, \quad t \geq 0$$

Ide o takzvaný **regulovaný Brownov pohyb** (*Reflected Brownian Motion*), nakoľko

- proces Z_t^j nadobúda len nezáporné hodnoty
- proces z_t^j nadobúda len nekladné hodnoty

Z_t^j a z_t^j má následne folded (pre $Z_0^j = 0$ polo) normálne rozdelenie.

Podľa Itoovej lemy sa logaritmy cien $Y_t^j = \ln(X_t^j)$ vyvíjajú ako Brownov pohyb

$$dY_t^j = \mu_j dt + \sigma_j dW_t^j, \quad j = 1, 2.$$

Definujme procesy

$$Z_t^j = Y_t^j - L_t^j = Y_t^j - \inf_{0 \leq s \leq t} Y_s^j \quad \text{a} \quad z_t^j = Y_t^j - H_t^j = Y_t^j - \sup_{0 \leq s \leq t} Y_s^j, \quad t \geq 0$$

Ide o takzvaný **regulovaný Brownov pohyb** (*Reflected Brownian Motion*), nakoľko

- proces Z_t^j nadobúda len nezáporné hodnoty
- proces z_t^j nadobúda len nekladné hodnoty

Z_t^j a z_t^j má následne folded (pre $Z_0^j = 0$ polo) normálne rozdelenie.

Podľa Itoovej lemy sa logaritmy cien $Y_t^j = \ln(X_t^j)$ vyvíjajú ako Brownov pohyb

$$dY_t^j = \mu_j dt + \sigma_j dW_t^j, \quad j = 1, 2.$$

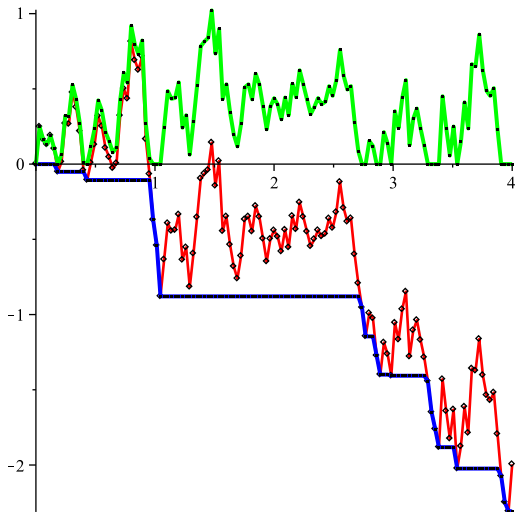
Definujme procesy

$$Z_t^j = Y_t^j - L_t^j = Y_t^j - \inf_{0 \leq s \leq t} Y_s^j \quad \text{a} \quad z_t^j = Y_t^j - H_t^j = Y_t^j - \sup_{0 \leq s \leq t} Y_s^j, \quad t \geq 0$$

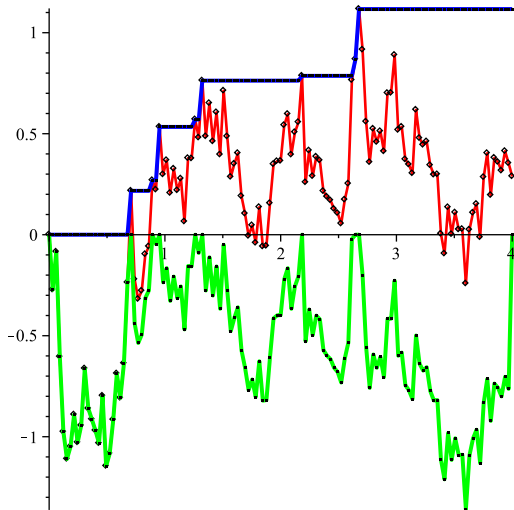
Ide o takzvaný **regulovaný Brownov pohyb** (*Reflected Brownian Motion*), nakoľko

- proces Z_t^j nadobúda len nezáporné hodnoty
- proces z_t^j nadobúda len nekladné hodnoty

Z_t^j a z_t^j má následne folded (pre $Z_0^j = 0$ polo) normálne rozdelenie.

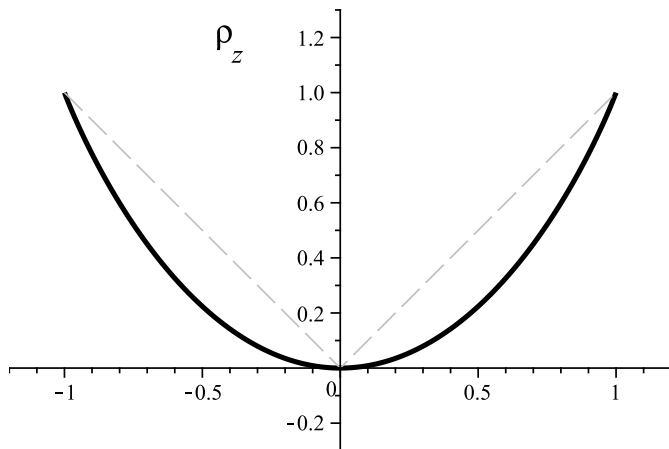


Typický priebeh W_t (červená), L_t (modrá) a Z_t (zelená).



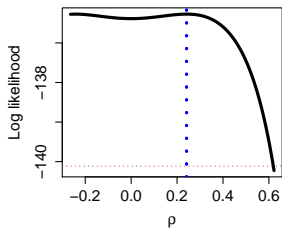
Typický priebeh W_t (červená), H_t (modrá) a z_t (zelená).

Vzťah medzi koreláciami ρ a ρ_z

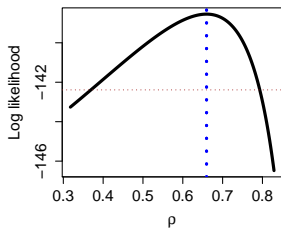


MLE korelácií pri half-normal rozdelení

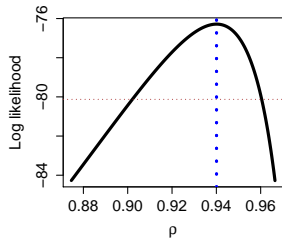
Estimated value, rho = 0.207



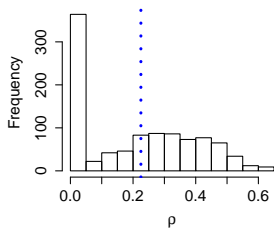
Estimated value, rho = 0.675



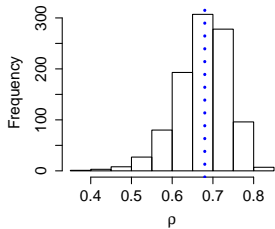
Estimated value, rho = 0.929



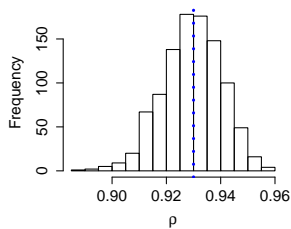
real value, rho = 0.225



real value, rho = 0.68



real value, rho = 0.93



Výsledky

- Využitie **úplnejšej informácie** – OHLC ceny namiesto uzatváracích.
- Lepšie asymptotické vlastnosti **estimátora** pre niektoré intervaly korelácií.
- Odvodené všetko potrebné pre odhad maximum likelihood.

Ďalšie smerovanie

- Empirická analýza.
- Vlastnosti pre nízke korelácie.
- Cenové skoky, napr. zatváracie a otváracie ceny.

Estimating correlation from OHLC prices

Jozef Kiseľák – Štefan Lyócsa – Tomáš Výrost

Analytické metódy a Big Data manažment v praxi
7.12.2015

Acknowledgements

Autori by chceli poďakovať za podporu tejto prezentácie z prostriedkov grantu Nadácie VUB číslo 2015-3-02/5.

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0666-11 and No. APVV-14-0357.

Authors also appreciate the support provided from the Slovak Grant Agency for Science (VEGA project No. 1/0392/15).



SLOVAK RESEARCH
AND DEVELOPMENT
AGENCY