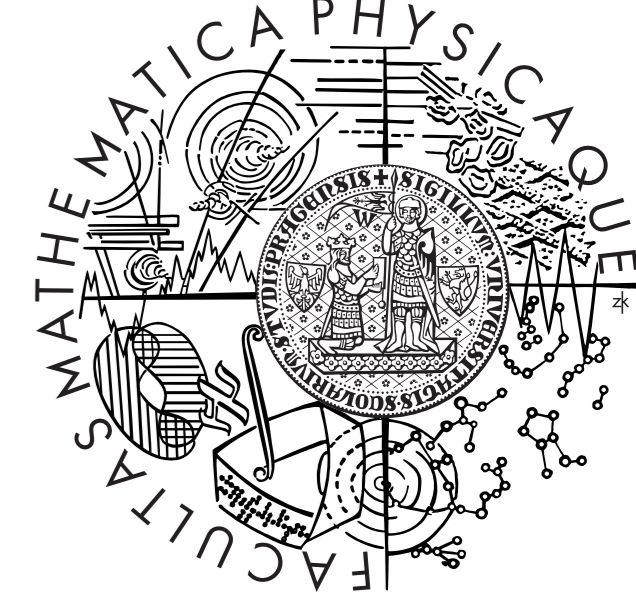




Funkcionální podmíněné kvantily rizikově neutrálních hustot

Zdeněk Hlávka

hlavka@karlin.mff.cuni.cz
Univerzita Karlova v Praze, KPMS MFF UK, Sokolovská 83, Praha



Kupní cenu $C_t(T)$ opce evropského typu v čase t s maturitou $T = t + \tau$, $\tau > 0$, lze zapsat jako střední hodnotu zisku $z(S_T)$, závisícího na budoucí ceně S_T podkladového instrumentu v čase T , opravenou o bezrizikovou úrokovou míru r : $C_t(T) = \exp\{-r\tau\}Ez(S_T) = \exp\{-r\tau\} \int_0^{+\infty} z(S_T) f_t(S_T) dS_T$, kde rizikově neutrální hustotu $f_t(\cdot)$ lze interpretovat jako pravděpodobnostní hustotu náhodné veličiny S_T . Odhad rizikově neutrální hustoty bývá obvykle odvozen od cen opcí [5]. Cílem příspěvku je porovnání skutečných budoucích cen $S_{t+\nu}$, $\nu > 0$, s různými předpověďmi založenými na funkcionální neparаметrické regresi i na podmíněných funkcionálních kvantilech.

1 Rizikově neutrální hustota (RNH)

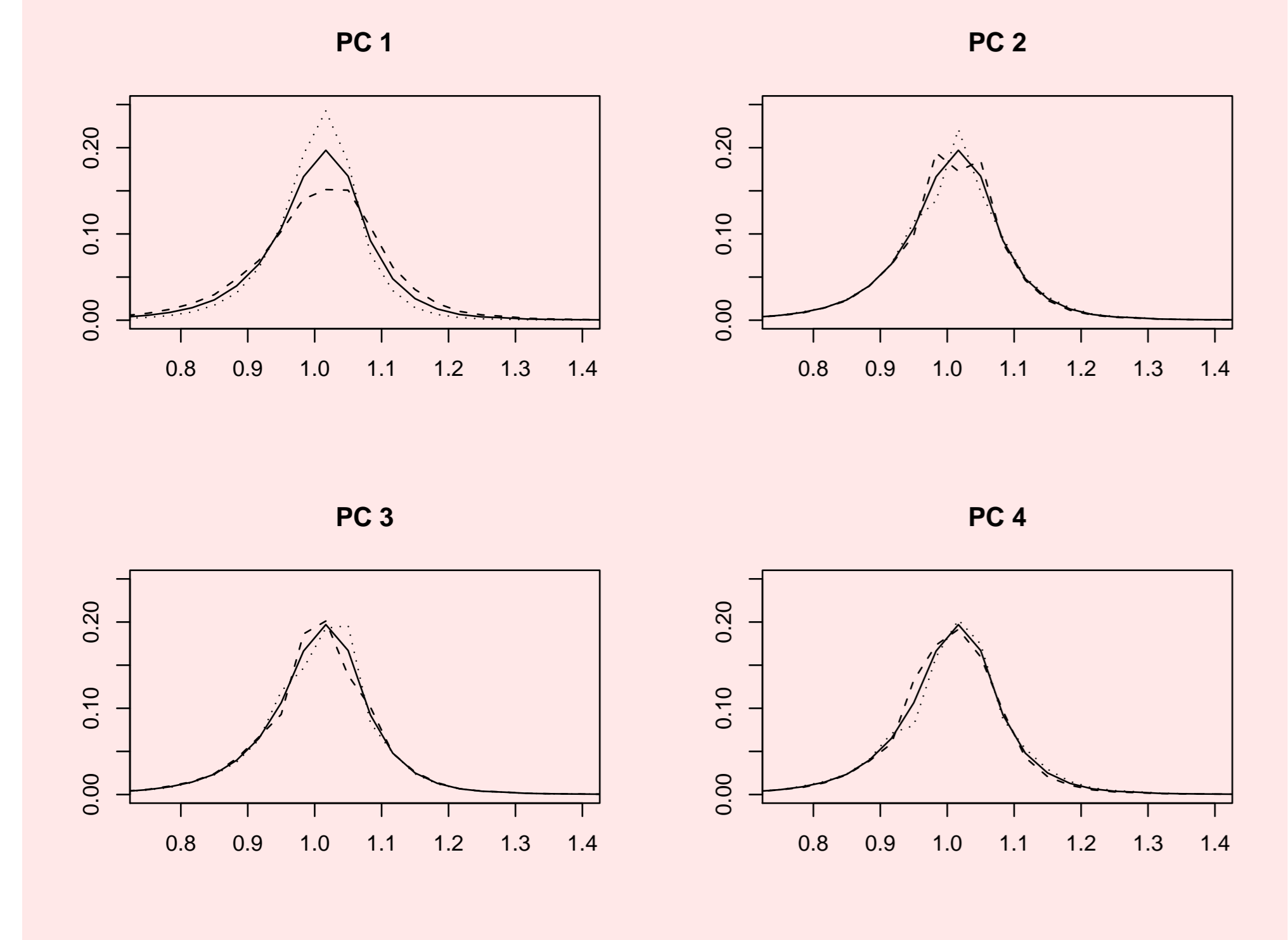
Majitel kupní opce evropského typu získá v čase T částku $z(S_T) = (S_T - K)_+ = \max(S_T - K, 0)$. Cena této opce v čase t je:

$$C_t(K, T) = \exp\{-r(T-t)\} \int_0^{+\infty} (S_T - K)_+ f_t(S_T) dS_T.$$

Pro odhad RNH $f_t(\cdot)$ je důležité, že v tomto případě ji lze vyjádřit jako druhou derivaci ceny opce vzhledem k realizační ceně K [2]:

$$f_t(K) = \exp\{r(T-t)\} \frac{\partial^2 C_t(K, T)}{\partial K^2}.$$

V článku [4] je ukázáno, že funkcionální hlavní komponenty [6] RNH lze interpretovat jako změny tvaru pozorovaných funkcí:



2 Neparаметrická funkcionální regrese

Funkcionální náhodná proměnná \mathcal{X} je v [3] obecně zavedena jako měřitelné zobrazení z pravděpodobnostního prostoru $(\Omega, \mathcal{A}, \mathcal{P})$ do nekonečné rozměrného měřitelného prostoru (G, \mathcal{G}) .

Za závisle proměnnou zvolme náhodnou proměnnou $S_{t+\nu}$, $\nu > 0$, a za vysvětlující funkcionální proměnnou odhad RNH $\hat{f}_t(\cdot)$, kde $t = 1, \dots, n$.

Funkcionální jádrové neparаметrické odhady zavedeme jako vážené lokální průměry závisle proměnné, zde vyjádřené pro předpověď ceny odpovídající $\hat{f}_{n+1}(\cdot)$:

$$\hat{S}_{n+1+\nu} = r\{\hat{f}_{n+1}(\cdot)\} = \frac{\sum_{t=1}^n S_{t+\nu} K_h[d\{\hat{f}_{n+1}(\cdot), \hat{f}_t(\cdot)\}]}{\sum_{t=1}^n K_h[d\{\hat{f}_{n+1}(\cdot), \hat{f}_t(\cdot)\}]}, \quad (1)$$

kde $K_h(\cdot)$ je asymetrická jádrová funkce s parametrem h a $d(\cdot)$ je vhodná míra vzdálenosti dvou funkcionálních pozorování [3].

Volba jádrové funkce ve vzorci (1) je obdobná jako u obyčejné jádrové regrese a rozdíl je pouze v tom, že kvůli použití vzdálenosti uvnitř jádrové funkce se ve vzorci (1) musí používat asymetrické jádrové funkce, tj. jádrové funkce, které jsou nenulové pouze pro nezáporné hodnoty vzdálenosti. Volba parametru h jádrové funkce je v praxi obvykle vyřešena pomocí tzv. křížového ověřování (crossvalidation), podrobněji viz sekce 3.

Mnohem komplikovanější je při analýze funkcionálních dat volba vzdálenosti. V práci [3] se místo s Euklidovskou vzdáleností pracuje s tzv. semimetrikou: $d(\cdot)$ je semimetrika na prostoru G , pokud pro každé $x \in G$ platí $d(x, x) = 0$ a zároveň pro každou trojici $(x, y, z) \in G^3$ platí $d(x, y) \leq d(x, z) + d(z, y)$. V praxi se doporučují semimetricky založené na hlavních komponentách (PCA), partial least squares nebo na derivacích. Volba semimetricky závisí na situaci a na typu a vlastnostech analyzovaných funkcionálních dat. My navážeme na analýzu funkcionálních hlavních komponent z článku [4] a budeme používat semimetricku založenou na hlavních komponentách.

Semimetrika typu PCA: Jsou-li pozorované funkce definovány svými hodnotami na mřížce u_1, \dots, u_m , je metoda funkcionálních hlavních komponent [6] ekvivalentní obyčejné metodě hlavních komponent: pomocí spektrálního rozkladu $V_n = \Gamma \Gamma^T$ varianční matice V_n vektorů $\hat{f}_t = (\hat{f}_t(u_1), \dots, \hat{f}_t(u_m))^T$, $t = 1, \dots, n$, získáme funkcionální hlavní komponenty $\xi_t = \hat{f}_t \Gamma$ a datovou matici $F(n \times m)$ s řádky \hat{f}_t , $t = 1, \dots, n$, můžeme zapsat jako $F = (F \Gamma) \Gamma^T$. Označíme-li symbolem Γ_q matici obsahující prvních q sloupců matice Γ , získáme v jistém smyslu optimální aproximaci $F_q = (F \Gamma_q) \Gamma_q^T$, kterou použijeme k zavedení semimetricky (vzdálenosti) typu PCA mezi diskretizovanými funkcionálními pozorováními $x = (x(u_1), \dots, x(u_m))^T$ a $y = (y(u_1), \dots, y(u_m))^T$:

$$d_q^{PCA}(x, y) = \|x - y\|_q = \sqrt{(x - y)^T \Gamma_q \Gamma_q^T \{(x - y)^T \Gamma_q \Gamma_q^T\}^T} \\ = \sqrt{\sum_{k=1}^q \{(x - y)^T \gamma_k\}^2} = \sqrt{\sum_{k=1}^q \left[\sum_{i=1}^m \{x(u_i) - y(u_i)\} \gamma_{ik} \right]^2}.$$

Pro funkcionální pozorování $x(\cdot)$ a $y(\cdot)$ můžeme semimetricku typu PCA zapsat jako:

$$d_q^{PCA}\{x(\cdot), y(\cdot)\} = \sqrt{\sum_{k=1}^q \left[\int \{x(t) - y(t)\} \gamma_k(t) dt \right]^2},$$

kde $\gamma_k(\cdot)$ jsou příslušné vlastní funkce.

3 Implementace v R

	specifické parametry
<code>funopare.kernel()</code>	<code>bandwidth</code> (nemá defaultní nastavení)
<code>funopare.kernel.cv()</code>	
<code>funopare.knn()</code>	<code>neighbour</code> (nemá defaultní nastavení)
<code>funopare.knn.gcv()</code>	
<code>funopare.knn.lcv()</code>	
<code>funopare.mode.lcv()</code>	<code>Knearest=NULL</code>
<code>funopare.quantile.lcv()</code>	<code>Knearest=NULL, alpha=c(.05, .5, .95)</code>

Všechny výše uvedené funkce pro neparаметrickou funkcionální regresi ([3], <http://www.lsp.ups-tlse.fr/staph/npfda/>) požadují závisle proměnnou (**Response**), nezávisle proměnnou (**CURVES**) a funkce, pro které se mají spočítat předpovězené hodnoty (**PRED**). Všechny funkce umožňují zvolit jádrovou funkci (**kind.of.kernel**) a semimetricku (**semimetric**).

Následující funkce počítají odhad (1) a liší se pouze volbou parametru h :

funopare.kernel() Hodnota h je nastavena parametrem `bandwidth`.

funopare.kernel.cv() Hodnota h je určena pomocí křížového ověřování.

funopare.knn() Parametr `neighbour` určuje počet nejbližších pozorování, která budou použita k výpočtu v každém bodě. Hodnota parametru h_t , $t = 1, \dots, n$, je pak pro každé pozorování zvolena tak, aby do tohoto okolí padl právě zadaný počet "nejbližších sousedů".

funopare.knn.gcv() Počet nejbližších sousedů určující lokální šířku okénka h_t je určen globálně pomocí křížového ověřování.

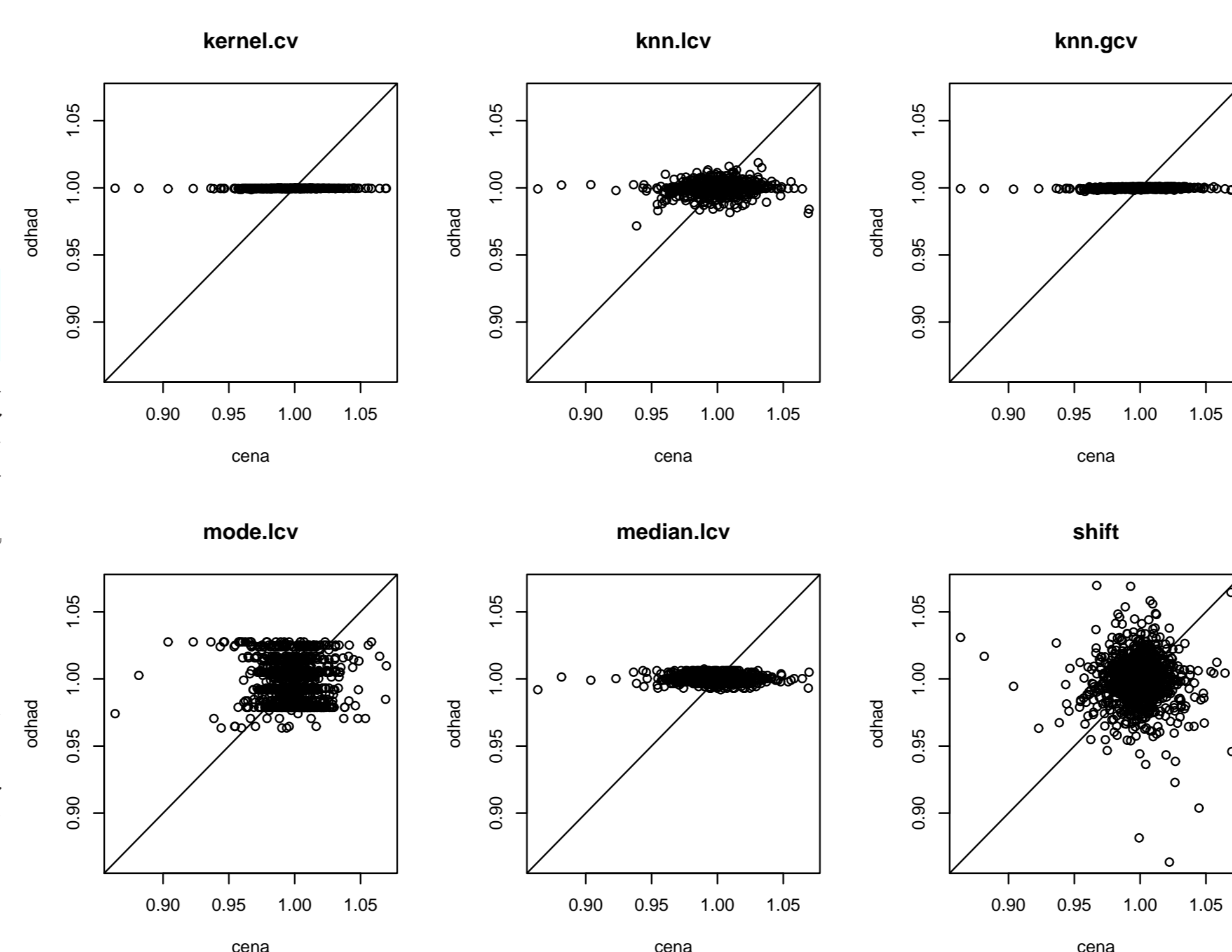
funopare.knn.lcv() Počet nejbližších sousedů je v každém bodě určen zvlášť pomocí křížového ověřování.

Zbývající dvě funkce jsou založené na jádrových odhadech podmíněné distribuční funkce a podmíněné hustoty: funkce **funopare.mode.lcv()** počítá podmíněný modus a **funopare.quantile.lcv()** s parametrem `alpha=0.5` počítá podmíněný medián.

4 Neparаметrická funkcionální regrese na odhadech RNH

Neparаметrickou funkcionální regresi použijeme na RNH odhadnuté z cen opcí na index DAX od ledna 1995 do března 2003. Pro odstranění závislosti odhadnutých hustot na aktuální hodnotě DAXu jsou odhady RNH spočítány v bodech odpovídajících násobkům současné hodnoty DAXu; interval od 0 do 4-násobku současné hodnoty je rovnoměrně rozdělen na 120 částí a každá "funkce" je tak diskretizována pomocí 121 funkčních hodnot. Pomocí neparаметrické funkcionální regrese se nyní pokusíme vyšetřit vztah mezi cenami akcií S_{t+1} a odhady rizikově neutrálních hustot $f_t(\cdot)$. Parametr semimetricky PCA byl zvolen $q = 20$.

Data jsme rozdělili na dvě části. První část dat slouží pouze pro odhad neparаметrické funkcionální regrese a na druhé části dat jsme takto získaný odhad vyzkoušeli. Předpovědi spočítané v testovací části našich dat všemi metodami s automatickou volbou parametru h jsou znázorněny na obrázku (pro srovnání je v pravém dolním rohu i předpověď založená pouze na předchozím pozorování):



V tabulce jsou uvedeny střední čtvercové chyby ($\times 10^4$) v prvním a ve druhém (testovacím) výběru:

	kernel	knn.lcv	knn.gcv	mode	median	shift	constant
první	1.50	1.21	1.50	0.97	1.15	2.79	1.49
druhý	3.17	3.22	3.14	5.52	3.29	6.56	3.17

Je zajímavé, že v prvním řádku je nejlepší odhad založený na modu podmíněné hustoty, který zcela selhává na testovacích datech. Pro srovnání jsou uvedeny i hodnoty střední čtvercové chyby pro odhad založený pouze na předchozím pozorování (shift) a pro konstantní odhad rovný průměrné ceně (constant). Odhad založený na předchozím pozorování je jednoznačně nejhorší, zatímco konstantní odhad patří na testovacích datech k těm lepším. Výsledky této analýzy lze stručně shrnout očekávaným závěrem, že tvar odhadu rizikově neutrální hustoty neumožňuje předpovídat budoucí vývoj ceny podkladového instrumentu. Vhodnější a zajímavější zde budou neparаметrické funkcionální kvantily, které místo bodové předpovědi poskytují předpovědní intervaly pro budoucí hodnoty ceny DAXu.

5 Funkcionální podmíněné kvantily

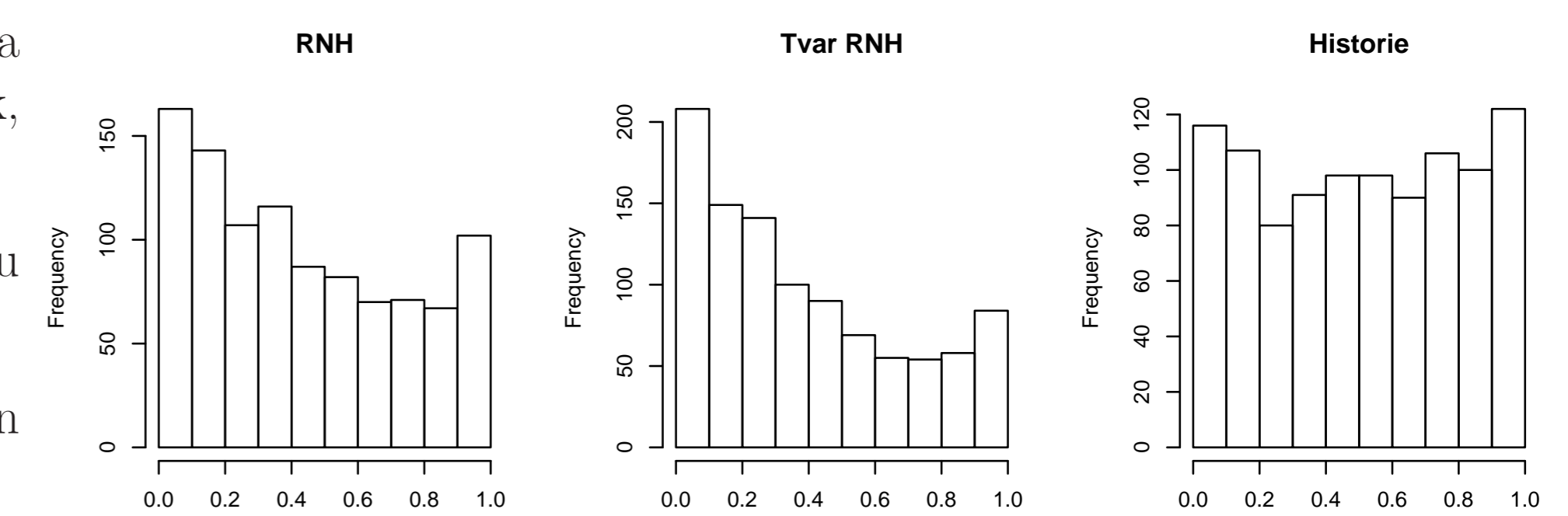
Odhad RNH $\hat{f}_t(\cdot)$ lze za předpokladu rizikové neutrality interpretovat jako odhad pravděpodobnostní hustoty náhodné veličiny $S_{t+\tau} = S_T$. Ve skutečnosti se však v odhadech RNH projevuje averze k riziku, kterou lze jednoduše ukázat např. porovnáním kvantilů $\hat{f}_t(\cdot)$ s pozorovanými hodnotami S_T . V této části srovnáme kvantily odhadů RNH $\hat{f}_t(\cdot)$ s neparаметrickými funkcionálními kvantily založenými na tvaru $\hat{f}_t(\cdot)$ a s neparаметrickými funkcionálními kvantily založenými na historických pozorováních S_u , $u < t$. Soustředíme se na následující tři odhady kvantilů:

RNH Kvantil u_α^{RNH} je řešením rovnice $\int_{-\infty}^{u_\alpha^{RNH}} \hat{f}_t(x) dx = \alpha$.

Tvar RNH Kvantil u_α^{tvar} je určen z neparаметrického funkcionálního odhadu podmíněné distribuční funkce náhodné veličiny S_T . Vysvětlující funkcionální proměnná je zde odhad RNH $\hat{f}_t(\cdot)$ a volba semimetricky PCA zaručuje, že regresní funkce bude záviset na tvaru RNH.

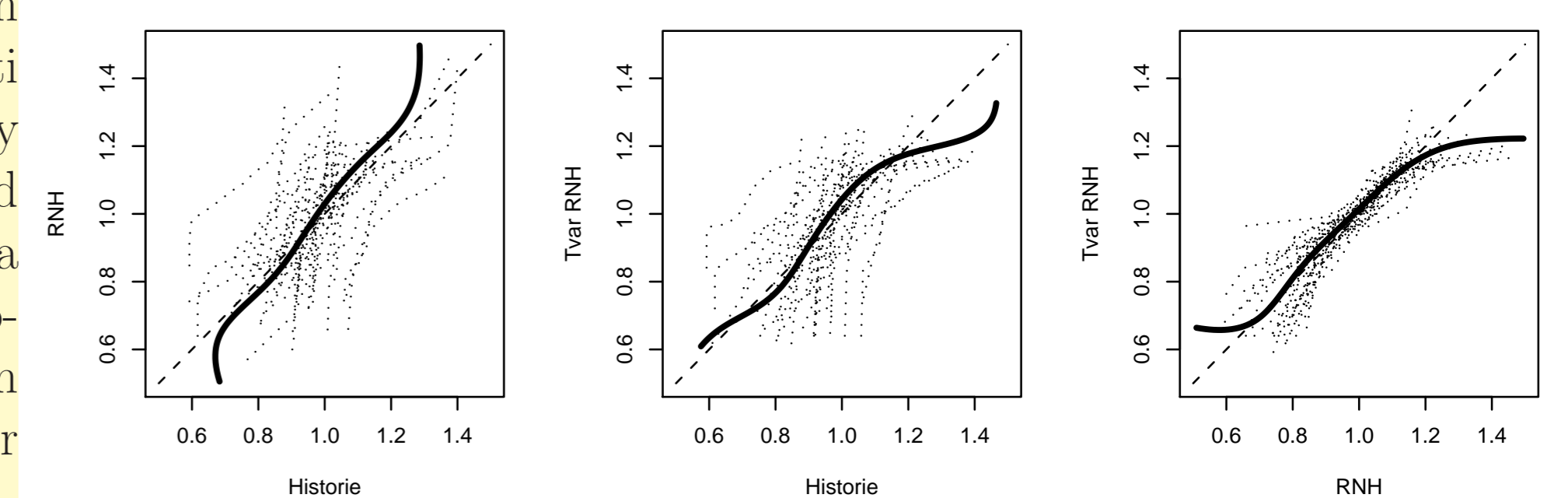
Historie Neparаметrické funkcionální kvantily u_α^{hist} s funkcionálními vysvětlujícími pozorováními danými historickými hodnotami S_u , $u < t$.

Na obrázku jsou nakresleny histogramy hodnot odhadů distribučních funkcí odpovídajících výše uvedeným kvantilům v bodech S_T :



Pro dobře kalibrované předpovědi by tyto histogramy měly být co nejpodobnější rovnoměrnému rozdělení (označíme-li skutečnou distribuční funkci cen S_T jako F_T , pak má náhodná veličina $F_T(S_T)$ rovnoměrné rozdělení). Vidíme tedy, že nejrealističtější předpovědi poskytuje neparаметrická funkcionální regrese založená na historických cenách, zatímco u zbylých metod příliš často nastávají události odpovídající nízkým kvantilům.

Následující kvantilové diagramy umožňují přímé srovnání získaných kvantilů:



Pro přehlednost je tečkovanou čarou znázorněna jenom menší část kvantilů (každý pětadvacátý). Silná čára označuje lokálně lineární odhad spočítaný ze všech kvantilových diagramů (kvůli symetrii spočítaný z dat otočených o $\pi/4$). Srovnání ukazuje, že kvantily RNH často předpovídají příliš nízké nebo příliš vysoké hodnoty, zatímco neparаметrická funkcionální regrese založená na tvaru RNH je příliš konzervativní pro vyšší kvantily.

6 Závěr

V části 4 jsme neparаметrické funkcionální metody použili pro bodovou předpověď ceny a ověřili jsme, že cenu podkladového instrumentu nelze touto metodou spolehlivě předpovídat. V části 5 jsme pomocí neparаметrických funkcionálních kvantilů zkonstruovali rozmezí, ve kterém by se cena podkladového instrumentu měla v budoucnu nacházet a porovnáním různých typů kvantilů jsme se pokusili názorně popsat projevy averze k riziku, která je přítomná v cenách opcí a tedy i v odhadech RNH [1].

Nevýhodou neparаметrických funkcionálních metod je obtížná interpretace výsledků a nemožnost přesvědčivého grafického znázornění výsledků. Proto můžeme neparаметrické funkcionální metody doporučit spíše pro průzkumovou analýzu dat, než pro detailní vyšetřování regresních závislostí.

Reference

- [1] Ait-Sahalia, Y., Wang, Y., Yared, F. (2000) *Do option markets correctly price the probabilities of movement of the underlying asset?*. Journal of Econometrics **102**, 67–110.
- [2] Breeden, D., Litzenberger, R. (1978) *Prices of state-contingent claims implicit in option prices*. Journal of Business **51**, 621–651.
- [3] Ferraty, F., Vieu, P. (2006) *Nonparametric Functional Data Analysis*. Springer, New York.
- [4] Hlávka, Z. (2006) *Funkcionální hlavní komponenty pro odhady RNH*, Sborník prací 13. letní školy JČMF Robust 2006, 91–98.
- [5] Jackwerth, J.C. (1999) *Option-implied risk-neutral distributions and implied binomial trees: a literature review*. Journal of Derivatives **7**, 66–82.
- [6] Ramsay, J.O., Silverman, B.W. (1997) *Functional Data Analysis*. Springer, New York.

Poděkování Tato práce byla podporována grantem GAČR 201/08/0486 a výzkumným záměrem MSM 0021620839.