Současná ekonometrie jako součást matematiky

Zuzana Prášková, Tomáš Cipra, Jitka Dupačová a Petr Lachout, Praha

1. Úvod

Ekonometrie je oblast matematiky zabývající se jejími aplikacemi v ekonomii. Autoři učebnic o ekonometrii s oblibou uvádějí, že tuto disciplínu lze vymezit jako průnik matematické ekonomie, matematické statistiky a ekonomické statistiky, navíc se zde stále zřetelněji uplatňuje matematické modelování ekonomických jevů a procesů. Pro modelování ekonomických zákonitostí, předpovídání budoucích vývojových trendů v podmínkách neúplné a nepřesné informace a pro podporu rozhodování v rámci složitých dynamických ekonomických systémů je možné a potřebné používat matematické metody, a to především metody pravděpodobnosti a matematické statistiky, ale také např. nestandardní optimalizační metody, diferenciální a diferenční rovnice, maticový počet, metody numerické matematiky atd. Konkrétní aplikace jsou nemyslitelné bez využití výpočetní techniky a speciálního statistického nebo optimalizačního programového vybavení.

Jestliže pomineme nové ekonomické teorie z počátku 19. století (např. analýzu nabídky a poptávky, rozdělení důchodů, teorii produkčních funkcí) a vliv rozvoje statistiky na počátku 20. století, pak je moderní ekonometrie spojena až se založením Ekonometrické společnosti 29. prosince 1930 a s vydáváním časopisu Econometrica. Stručná charakteristika Ekonometrické společnosti, tak jak ji najdeme na obalu časopisu Econometrica, říká, že jde o mezinárodní společnost pro rozvoj ekonomických teorií ve vztahu ke statistice a matematice.

Nosným tématem byl po dlouhou dobu rozvoj teorie lineárních soustav simultánních rovnic spolu s jejich aplikacemi podmíněnými dosaženým stupněm rozvoje výpočetní techniky; připomeňme např. Tinbergenův model holandské ekonomiky s 22 rovnicemi z roku 1937 a pozdější model britské ekonomiky zvaný „Cambridge-2" s 2552 rovnicemi z roku 1976. Současná ekonometrie reaguje na nově vznikající ekonomické jevy a potřeby (inflace, měnová problematika, rozvoj finančních trhů, privatizace celých ekono-

---

Doc. RNDr. Zuzana Prášková, CSc. (1946), prof. RNDr. Tomáš Cipra, DrSc. (1952), prof. RNDr. Jitka Dupačová, DrSc. (1939), a RNDr. Petr Lachout, CSc. (1958), jsou členy oddělení ekonometrie na katedře pravděpodobnosti a matematické statistiky, MFF UK, Sokolovská 83, 186 75 Praha 8.

Práce vznikla v rámci výzkumného záměru MŠMT Matematické metody ve stochastice (MSM 113200008) a jednotliví autoři byli podporováni také z prostředků Grantové agentury ČR v rámci grantů 201/00/0769, 201/00/0770, 401/99/1136 a 201/99/0264.

mik, nové přístupy v energetice, problémy ekonomik s ohledem na stárnoucí populaci atp.). Vzhledem k náročnosti a závažnosti zmíněných problémů význam ekonometrie v dnešní společnosti roste. Ekonometrické studie jsou používány např. jako podpůrný prostředek pro klíčová rozhodnutí a vymezení dalších strategií ekonomického rozvoje. V této oblasti není možné konat opakované experimenty nebo čekat, až bude dostatek homogenních pozorování. Experimentování i odklady rozhodnutí by měly nedozírné následky.

Atraktivními oblastmi výzkumu jsou např. metody pro analýzu a řízení složitých rozhodovacích procesů včetně vyhodnocení získaných výsledků, generování scénářů vývoje úrokové míry pro finanční modely, řízení účelových portfolií s ohledem na vyvíjející se makroekonomické i mikroekonomické prostředí, analýza časových řad s chybějícími pozorováními a se složitou strukturou reziduí, nestacionární modely, modely s kvalitativními, průřezovými nebo panelovými daty atp. Jejich společným rysem je snaha odstranit nereálné předpoklady stávajících modelů a postupů a tím se přiblížit skutečnosti. Z matematického hlediska tak vznikají velmi zajímavé problémy, které jsou přitom motivovány potřebami praxe. Místo soustav lineárních simultánních rovnic se studují a aplikují nelineární modely (Dhrymes, 1994b), modely s časově proměnnými parametry, pro odhady a testy se využívají modifikované postupy robustní a neparametrické statistiky (Koenker, 1982). Získané výsledky jsou podrobovány detailní analýze, která je často podnětem ke změně modelu, k dalšímu zjemnění či zpřesnění, k doplnění dat nebo dává informaci o přesnosti výsledků.

Dále budeme stručně ilustrovat vybrané problémové a metodické okruhy, kterými se současná ekonometrie zabývá. Podrobněji se budeme věnovat aktuálním trendům v modelování časových řad a metodě bootstrap, která má obecnější použití v analýze statistických dat. V závěru uvedeme příklad komplexního řešení reálného problému souvisejícího s evropskou ekonomickou integrací. Kromě citací prací, na něž se v textu odkazujeme, je seznam literatury rozšířen o několik učebnic a monografií věnovaných ekonometrii.

2. Dynamické modely a jejich stavová reprezentace

Modely tohoto typu se ukázaly jako velmi užitečné v oblasti dynamické ekonomie, v moderní teorii řízení a v analýze časových řad, neboť bez velkých výpočetních problémů pracují s časově proměnnými parametry. *Kalmanův filtr*, figurující v této souvislosti jako nejpoužívanější výpočetní nástroj, byl původně navržen pro různé inženýrské aplikace (např. pro navádění střel apod., viz Kalman (1960)), později se však ukázaly jeho přednosti také v bayesovské statistice, v dynamických ekonometrických modelech či při konstrukci předpovědí v časových řadách.

V nejjednodušší podobě je možné Kalmanův filtr představit na příkladu dynamického lineárního modelu tvaru

$$X_t = G_t X_{t-1} + W_t, \tag{1}$$

$$Y_t = F_t X_t + V_t, \tag{2}$$

kde X_t je n-rozměrný *stavový vektor* popisující stav systému v čase t (např. vektor všech parametrů ekonometrického modelu, které je nutno odhadnout) a Y_t je m-rozměrný *vektor pozorování* v čase t; F_t a G_t jsou matice (neznámých) konstant. Vychází se zde z představy, že uvažovaný systém je plně popsán znalostí jeho stavu X_t, který se v nejjednodušším případě vyvíjí v čase pomocí *stavové rovnice* (1). Tento stav ovšem není pozorovatelný přímo, ale jen prostřednictvím pozorovaných hodnot Y_t souvisejících se stavovým vektorem díky *rovnici pozorování* (2). Obě uvedené rovnice mají stochastickou složku modelovanou pomocí náhodných veličin W_t a V_t.

V nejpoužívanější podobě poskytuje Kalmanův filtr rekurentní vzorce pro odhad (jistých optimálních vlastností) střední hodnoty stavu a chyby tohoto odhadu. Pomocí této metodiky a jejích zobecnění je možné také konstruovat předpovědi stavů i pozorovaných hodnot, odhadovat pravděpodobnostní stavové rozdělení, zabývat se nelineárními závislostmi, zohlednit chybějící a odlehlá pozorování apod.

Aktuální ekonometrické aplikace (spotřeba plynu ve Velké Británii, volatilita denních kurzů libry vůči dolaru) lze nalézt např. v práci Durbin a Koopman (2000).

3. Kointegrace ekonomických veličin

Dalším z důležitých pojmů v současné ekonometrii je pojem *kointegrace*, který se v analýze ekonomických dat objevuje od poloviny osmdesátých let (viz např. Engle a Granger (1987)). Velmi zjednodušeně řečeno je kointegrace určitý rovnovážný stav systému několika časových řad. Při konstrukci modelů ekonomických časových řad je totiž přirozené vycházet z následujícího předpokladu: pokud jsou jednotlivé řady svázány teoreticky zdůvodněným ekonomickým vztahem (např. diskontní sazba a inflace nebo peněžní zásoba, rozpočtové výdaje a výnosy státních dluhopisů), pak se jejich vývoj v dlouhodobém časovém horizontu nerozchází. Příslušné stabilní stavy se v této souvislosti označují jako stacionární.

Nechť např. časové řady $\{Y_t\}$ a $\{Z_t\}$ jsou obě typu $I(1)$, to znamená, že je lze stacionarizovat přechodem k řadě prvních diferencí $\{\Delta Y_t\}$ a $\{\Delta Z_t\}$, kde např. $\Delta Y_t = Y_t - Y_{t-1}$. Pak se tyto dvě řady nazývají kointegrované, jestliže existuje jejich lineární kombinace $\{aY_t + bZ_t\}$, která již je stacionární. Nalezení kointegračních vztahů mezi řadami ekonomických ukazatelů včetně odhadu lineárních koeficientů typu a a b (tzv. *kointegrační vektor*) má závažné důsledky pro modelování celého systému.

V souvislosti s kointegrací se objevuje celá řada dalších problémů, které jsou typické pro ekonomickou praxi, např. *problém kauzality* (je daný ekonomický jev příčinou jiného jevu, či nikoli?), *testování jednotkových kořenů* pomocí tzv. Dickeyova-Fullerova testu (Dickey a Fuller, 1979) a jeho četných modifikací (lze vykazované nestacionární trendové či periodické chování dynamického ekonomického jevu odstranit jednoduše přechodem k odpovídající řadě diferencí původní řady?), *zdánlivá regrese* (proč dvě řady, které spolu věcně naprosto nesouvisejí, vykazují významnou statistickou závislost?) apod.

Řešení těchto problémů je zpravidla spojeno s nalezením vhodných testových statistik a odhadů a studiem jejich pravděpodobnostního chování. Zde se objevuje řada matematických problémů, které je třeba řešit nestandardními postupy, neboť nestacionární procesy mají zcela jiné asymptotické vlastnosti, než vykazují procesy stacionární. Dalším problémem je nalezení lepších aproximací, než jaké poskytuje asymptotická teorie, vzhledem k tomu, že rychlost konvergence některých asymptotických postupů je pro praktické využití příliš pomalá.

Více informací o problémech kointegrace lze nalézt např. v Engle a Granger (1987, 1991) nebo v Banerjee et al. (1993).

4. Regresní modely s kategoriálními endogenními veličinami

Modely tohoto typu se často používají při hodnocení biologických pokusů; v současné době se však aplikují i v ekonometrii, neboť i v ekonomických datech často vystupují proměnné ryze kategoriální povahy, viz např. Fomby et al. (1984) nebo Kennedy (1998). Velmi častý je případ dichotomické proměnné, která nabývá pouze hodnot 0 a 1.

Jedním z modelů, kde se kategoriální data přirozeně vyskytují, je regrese s dichotomickou endogenní proměnnou. Obecně se uvažuje model nelineární regrese

$$Y_i = g(X_i, \theta) + \varepsilon_i, \quad i = 1, \ldots, n,$$

kde n je počet pozorování, která jsou k dispozici, a $\theta \in \Theta$ je parametr regresního modelu. K dispozici je dodatečná informace, že Y_i je buď 0 („jev nenastal“), nebo 1 („jev nastal“).

Uveďme si standardní příklad vzniku tohoto regresního vztahu. Uvažujme model prodeje nějaké komodity, který může být realizován pouze během n prodejních příležitostí; i-tá z nich je charakterizována úrovní nabídky (nebo užitku) U_i a nastavením podmínek prodeje $X_i \in \mathbb{R}^k$. Model má neznámý parametr $\beta \in \mathbb{R}^k$ a vlastní transakce probíhá tak, že k prodeji dojde pouze v tom případě, že $X_i'\beta > U_i$. Naším cílem je odhadnout parametr β, ale k dispozici jsou pouze údaje, že při i-té prodejní situaci k prodeji došlo či nedošlo, a jsou nastaveny podmínky X_i pro každou sledovanou prodejní příležitost.

Definujme náhodnou veličinu Y_i, pro kterou $Y_i = 1$, pokud při i-té prodejní situaci došlo k prodeji, a $Y_i = 0$, pokud se prodej neuskutečnil. Označme jako $P[A]$ pravděpodobnost jevu A a jako $I[A]$ indikátor množiny A. Můžeme tedy psát

$$Y_i = I[X_i'\beta > U_i] = P[X_i'\beta > U_i | X_i] + \varepsilon_i = F(X_i'\beta) + \varepsilon_i,$$

kde F je distribuční funkce nabídky U_i a ε_i je náhodná chyba.

V případě, že F je distribuční funkce standardního normálního rozdělení, mluvíme o *probitovém modelu*, v případě logistického rozdělení, tj. když $F(t) = \mathrm{e}^t/(1 + \mathrm{e}^t)$, o *logitovém modelu*. Odhad regresních koeficientů $\hat{\beta}_n$ se většinou v obou případech hledá metodou maximální věrohodnosti, což vede k řešení nekonvexní optimalizační úlohy bez omezení. Získané odhady jsou konzistentní s řádem konzistence $\sqrt{n}$ a normálním asymptotickým rozdělením, tj. náhodná veličina $\sqrt{n}(\hat{\beta}_n - \beta)$ má asymptoticky normální rozdělení s nulovou střední hodnotou a s konečným kladným rozptylem. Na tomto faktu je možné založit testy hypotéz, které jsou asymptoticky eficientní. Více o těchto modelech viz Fomby et al. (1984) nebo Kennedy (1998).

Problémem tohoto postupu je, že se předpokládá znalost distribuční funkce F. Jednou z možností, jak tento problém odstranit, je provést odhad ve třech stupních. Nejdříve se na základě předběžného odhadu parametru β, získaného například použitím probitového nebo logitového modelu, aproximují chyby ε_i. Tyto aproximace představují cenzorovaná data, na jejichž základě lze odhadnout distribuční funkci F, viz Andersen et al. (1993) nebo Fleming a Harrington (1984). Tohoto odhadu se pak použije pro nalezení konečného odhadu vektoru parametrů β.

5. Modely typu ARCH, GARCH

Konstrukce předpovědí budoucího vývoje ekonomických veličin z časové řady přítomných a minulých pozorování je jedním z nejdůležitějších problémů, kterými se současná ekonometrie zabývá. Velké oblibě se v 70. letech těšily modely časových řad typu ARMA (z anglického *autoregressive-moving average*), což jsou lineární modely, v nichž je současné pozorování X_t lineární kombinací předešlých pozorování X a přítomné a časově zpožděné šumové složky Y,

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + Y_t + \theta_1 Y_{t-1} + \theta_q Y_{t-q}, \tag{3}$$

později také integrované modely typu ARIMA, tj. modely nestacionárních časových řad, které lze převést na stacionární modely ARMA postupným diferencováním. Parametry těchto modelů lze snadno odhadnout a díky lineárním vazbám mezi veličinami lze snadno konstruovat předpovědi. Ukazuje se ale, že chybu předpovědi nelze zlepšit ani při dodatečné informaci o minulých pozorováních.

Mnoho ekonometrických časových řad, především finančních, však vykazuje nelineární dynamiku a zejména velmi silnou závislost okamžité variability na předchozím vývoji. Takové chování, které má vysokou míru nejistoty, nelze uspokojivě vysvětlit lineárními modely. Byly proto hledány nové postupy, jak modelovat časové řady.

V souvislosti se studiem inflace ve Velké Británii (Engle, 1982) byl navržen model, který uvažuje závislost chyby předpovědi na minulém průběhu časové řady a umožňuje lépe vysvětlit složitější strukturu náhodného procesu ekonomických veličin. Model byl nazván ARCH (*autoregressive conditional heteroscedasticity*), autoregresní model s podmíněnou heteroskedasticitou. V nejjednodušší verzi ARCH(1) je definován jako

$$\begin{aligned} X_t &= \sigma_t Y_t, \\ \sigma_t^2 &= a_0 + a_1 X_{t-1}^2, \end{aligned} \tag{4}$$

kde X_t je pozorování v čase t, Y_t je nezávislý šum s nulovou střední hodnotou a jednotkovým rozptylem a σ_t^2 je podmíněný rozptyl pozorování X_t při dodatečné informaci o pozorování v čase $t-1$, a_0, a_1 jsou konstanty, které musí splňovat určité podmínky vyplývající z požadavků stacionarity a kauzality procesu. V obecnější verzi se předpokládá, že σ_t^2 je (obecně nelineární) funkcí minulých pozorování až do času $t-p$. Model (4) lze ekvivalentně vyjádřit ve tvaru

$$X_t^2 = a_0 + a_1 X_{t-1}^2 + \eta_t, \tag{5}$$

kde $\eta_t = (Y_t^2 - 1)\sigma_t^2$ jsou nekorelované náhodné veličiny. Posloupnost $\{X_t^2\}$ má tedy formálně stejnou strukturu jako autoregresní posloupnost prvního řádu a parametry modelu mohou být odhadnuty standardními technikami užívanými v časových řadách; získané odhady však budou méně eficientní, neboť ignorují heteroskedasticitu chybové složky η_t. Při známém rozdělení chyb Y_t, resp. při známém podmíněném rozdělení veličin X_t, lze uplatnit metodu maximální věrohodnosti, která vede na nelineární optimalizační úlohu. Model byl původně odvozen za předpokladu gaussovského rozdělení chyb, empirické studie však ukazují, že finanční časové řady (řady cen opcí, vývoje směnných kurzů apod.) tento předpoklad nesplňují. Byly proto navrženy četné modifikace modelů ARCH, které zohledňují tuto skutečnost, zkoumány jejich vlastnosti, hledány robustní verze odhadů apod.

Skutečnost, že model (4) lze vyjádřit jako autoregresní model (5) druhých mocnin pozorování, vedla k přirozenému zobecnění modelů ARCH na modely, v nichž druhé mocniny pozorovaných veličin se řídí modelem ARMA, tj. modelem typu (3). Model GARCH (*generalized autoregressive conditional heteroscedasticity*) byl poprvé studován v článku Bollerslev (1986); jeho nejjednodušší verze GARCH$(1,1)$ je

$$\begin{aligned} X_t &= \sigma_t Y_t, \\ \sigma_t^2 &= a_0 + a_1 X_{t-1}^2 + b_1 \sigma_{t-1}^2, \end{aligned} \tag{6}$$

tedy připouští se, že podmíněný rozptyl pozorování v čase t, tj. σ_t^2, závisí nejen na pozorování v čase $t-1$, ale také na podmíněném rozptylu σ_{t-1}^2. Obecný lineární model GARCH(p,q) je formulován analogicky, předpokládá se, že σ_t^2 závisí na minulých hodnotách procesu X^2 až do času $t-q$ a na podmíněných rozptylech až do času $t-p$. Model (6) lze reprezentovat modelem ARMA$(1,1)$ ve tvaru

$$X_t^2 = a_0 + (a_1 + b_1) X_{t-1}^2 + \eta_t - b_1 \eta_{t-1},$$

odtud plynou podmínky na stacionaritu, kauzalitu a invertibilitu procesu. Reprezentace ARMA též umožňuje určit kovarianční strukturu procesu. Model GARCH$(1,1)$ se stal jedním z nejčastěji používaných modelů v analýze finančních časových řad, přestože bylo navrženo mnoho jiných modelů, které zobecňují a vylepšují jeho vlastnosti a které mají za cíl co nejlépe vysvětlit reálné děje probíhající na finančních trzích. Poznamenejme, že také studie Vošvrda et al. (1998), zabývající se analýzou indexu PX-50, tj. indexu 50 nejlikvidnějších titulů obchodovaných na Burze cenných papírů v Praze, vysvětluje proces $X_t = \log P_t - \log P_{t-1}$ denních změn indexu P_t jako autoregresní proces 1. řádu s chybami, které jsou typu GARCH$(1,1)$.

Zmiňme se ještě o modelech se *stochastickou volatilitou*, což jsou modely typu

$$\begin{aligned} X_t &= \sigma_t Y_t, \\ \log \sigma_t &= \alpha + \beta \log \sigma_{t-1} + \gamma Z_t, \end{aligned} \tag{7}$$

kde Y_t a Z_t jsou vzájemně nezávislé náhodné veličiny. V tomto modelu tedy podmíněný rozptyl X_t v čase t, resp. odpovídající směrodatná odchylka (volatilita), závisí ještě na náhodné veličině Z_t.

První rovnice v (7) se dá přepsat do tvaru

$$\log X_t^2 = \log \sigma_t^2 + \log Y_t^2,$$

což lze interpretovat jako rovnici pozorování, zatímco druhá rovnice v (7) je stavová rovnice v Kalmanově filtru popsaném v části 2. Pro odhad parametrů v modelu (7) a pro konstrukci předpovědí tedy lze užít metodiku vypracovanou pro Kalmanův filtr.

Řada zajímavých matematických problémů vzniká v souvislosti s aproximací modelů se stochastickou volatilitou a spojitým časem modely GARCH, se stacionaritou ARCH modelů nekonečného řádu či nejnověji s detekcí změn parametrů modelů ARCH a GARCH, viz např. Kokozska a Leipus (1998).

6. Bootstrap

Metoda *bootstrap*[1]) patří mezi tzv. intenzivní počítačové metody pro statistickou analýzu dat a její bouřlivý rozvoj v osmdesátých letech souvisí s vývojem nových počítačových technologií.

V krátké době poté, co byl publikován první článek o bootstrapu (Efron, 1979), byla zveřejněna celá řada dalších teoretických i simulačních studií, které měly za cíl zkoumat použití, účinnost a spolehlivost této metody v nejrůznějších aplikacích včetně ekonometrie a které zpětně přinesly i nové poznatky do asymptotické teorie empirických distribucí a procesů a nový pohled na použití metod Monte Carlo a generování náhodných čísel.

Základní problém je následující. Nechť $X_1, \ldots, X_n$ jsou nezávislé stejně rozdělené náhodné veličiny (náhodný výběr), jejichž pravděpodobnostní rozdělení F není blíže specifikováno. Nechť $\theta = \theta(F)$ je charakteristika rozdělení, která má určitou vypovídací hodnotu (např. průměr, směrodatná odchylka, medián); je to tedy neznámý parametr, který má být odhadnut na základě *realizace* náhodného výběru. Je-li stanovena výběrová statistika (funkce náhodných veličin) $T_n = T_n(X_1, \ldots, X_n)$ pro odhad parametru θ, je vhodné znát např. i vychýlení odhadu, tj. $E(T_n) - \theta$, kde $E(T_n)$ je střední hodnota počítaná jako

$$E(T_n) = \int T_n(y_1, \ldots, y_n)\, \mathrm{d}\big(F(y_1) \ldots F(y_n)\big), \tag{8}$$

a jeho směrodatnou odchylku $[E(T_n - E(T_n))^2]^{\frac{1}{2}}$; pro další statistické postupy (např. pro určení intervalů spolehlivosti nebo kritických hodnot pro testování hypotéz) je

[1]) Bootstrap, z anglického *to pull oneself up by one's bootstrap*, volně přeloženo *pomoz si sám.*

nutné určit rozdělení náhodné veličiny $R_n = R_n(X_1, \dots, X_n, F)$, což je vhodně standardizovaná statistika T_n (např. $R_n = \sqrt{n}(T_n - \theta)$) nebo nějaká její funkce. Je tedy třeba stanovit distribuční funkci

$$H_F(x) = P[R_n(X_1, \dots, X_n, F) \leqq x]. \tag{9}$$

Odvození přesného rozdělení $H_F(x)$ i výpočet vychýlení a dalších charakteristik mohou být v jednotlivých případech značně obtížné, nebo dokonce analyticky neproveditelné, a to i tehdy, když je distribuční funkce F známá. Je samozřejmě možné aproximovat H_F asymptotickým rozdělením, které lze odvodit na základě limitních vět teorie pravděpodobnosti. Přesnost takové aproximace však je ovlivněna a omezena počtem pozorování, která jsou skutečně k dispozici.

Metoda bootstrap kombinuje tzv. substituční princip a metodu Monte Carlo. Vychází z faktu, že empirická distribuční funkce F_n, tj.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I[X_i \leqq x],$$

při *skutečných hodnotách* $x_1, \dots, x_n$ veličin $X_1, \dots, X_n$ je *známá* funkce.

Jestliže potom $X_1^b, \dots, X_n^b$ je nezávislý náhodný výběr s vracením ze souboru pozorovaných hodnot $x_1, \dots, x_n$, znamená to, že $X_1^b, \dots, X_n^b$ jsou nezávislé stejně rozdělené náhodné veličiny, z nichž každá nabývá hodnot $x_1, \dots, x_n$ s pravděpodobností $1/n$ a má tedy distribuční funkci F_n. Soubor $X_1^b, \dots, X_n^b$ se nazývá *bootstrapový výběr*.

V dalších úvahách se původní výběr nahradí bootstrapovým výběrem a neznámá distribuce F známou distribucí F_n. Dostane se parametr $\theta^b = \theta(F_n)$ a statistiky $T_n^b = T_n(X_1^b, \dots, X_n^b)$ a $R_n^b = R_n(X_1^b, \dots, X_n^b, F_n)$. Potom např. vychýlení bootstrapového parametru θ^b lze vyjádřit jako $E^b T_n^b - \theta^b$, kde

$$E^b T_n^b = \int T_n(y_1, \dots, y_n) \, \mathrm{d}\big(F_n(y_1) \dots F_n(y_n)\big) \tag{10}$$

a rozdělení statistiky R_n^b jako

$$H_{F_n}(x) = P[R_n(X_1^b, \dots, X_n^b, F_n) \leqq x]. \tag{11}$$

Výpočty podle vzorců (10) a (11) jsou pro praktické použití vhodné jen v případě, že jejich pravé strany jsou explicitními funkcemi pozorování $X_1, \dots, X_n$. Stanovení výběrového rozdělení (11) by vyžadovalo provést všech n^n možných výběrů s vracením z populace $x_1, \dots, x_n$. To však lze uskutečnit jen pro výběry malého rozsahu (již pro 10 pozorování je 10 miliard možných výběrů). Proto se na bootstrapový výběr $X_1^b, \dots, X_n^b$ a známou distribuční funkci F_n aplikuje metoda Monte Carlo, kdy se mnohokrát (při rozsahu $n \approx 100$ se doporučuje řádově 10^3 až 10^4 opakování) generuje nezávislý náhodný výběr z rozdělení F_n, tj. mnohokrát se opakuje nezávislý náhodný výběr s vracením z populace čísel $x_1, \dots, x_n$, a při každém opakování se spočtou hodnoty T_n^b a R_n^b. Potom se např. distribuční funkce (11) a střední hodnota (10) spočtou jako empirická distribuční funkce a aritmetický průměr z hodnot, které získáme z dlouhé řady opakování tohoto umělého experimentu. Dostanou se tak *bootstrapové odhady* původního rozdělení a původních charakteristik.

Bootstrapové rozdělení (11) konzistentně odhaduje rozdělení $H_F(x)$, jestliže

$$\varrho(H_{F_n}, H_F) \to 0 \quad \text{při} \quad n \to \infty$$

v pravděpodobnosti nebo skoro jistě, kde ϱ je nějaká metrika na prostoru distribučních funkcí.

Konzistence metody bootstrap byla dokázána pro širokou škálu výběrových statistik; byla též studována rychlost konvergence bootstrapového rozdělení a asymptotika vyšších řádů. Teoretické i simulační studie se zabývají problémem volby rozsahu bootstrapového výběru a počtu opakování simulačního experimentu. Byly také navrženy různé modifikace této metody pro nestejně rozdělená nebo závislá pozorování, se kterými se v ekonometrii pracuje velmi často, modifikace, které zohledňují částečnou informaci o modelu, či naopak využívají informace vně modelu. Byly např. studovány bootstrapové odhady parametrů v dynamických lineárních modelech s exogenními proměnnými, v regresních modelech s heteroskedasticitami, odhady volatilit, byly navrženy bootstrapové verze Dickeyovy a Fullerovy statistiky a jejich modifikací pro testování jednotkových kořenů v modelech časových řad a dokázána konzistence takových postupů atd. Bootstrapové odhady lze také použít jako počáteční hodnoty ve složitých výpočetních procedurách, při konstrukci předpovědí při znalosti relativně malého počtu historických dat apod.

Další podrobnosti o metodě bootstrap lze nalézt např. v Efron a Tibshirani (1993) nebo v Shao a Tu (1995). Praktické využití přináší např. práce Abaffy et al. (1999).

7. Analýza postupu Evropské integrace

Termín ustavení Evropské měnové unie (EMU) byl významným podnětem pro ekonomickou a finanční integraci zainteresovaných zemí. Jeho fáze lze kvantifikovat různými způsoby. Práce Abaffy et al. (1999) analyzovala vliv jednotlivých členských zemí na výnosovou křivku vládních obligací celého EURO-bloku. Nebylo možné použít klasické metody založené na parametrické regresi, protože nebyly splněny předpoklady normality, homoskedasticity atp. Postupy založené na dvojstupňové neparametrické regresi však přinesly zajímavé výsledky.

Vstupní data jsou soubory výnosů do splatnosti $Y^j(t)$ vládních obligací zemí EMU (index j) pro obligace obchodované vždy v pevném dni; jde tedy o *průřezová data.* Předpokládáme, že výnos $Y^j(t)$ pro vládní obligaci s dobou do splatnosti t ve státě j má obecnou složku $m_1(t)$ a složku $m_2^j(t)$ specifickou pro stát j

$$Y^j(t) = m_1(t) + m_2^j(t) + \text{chyba},$$

kde m_1, m_2^j jsou hladké funkce, chyby jsou nezávislé, s nulovou střední hodnotou a v rovnici pro stát j mají rozptyl σ_j^2 nezávislý na době do splatnosti t. Funkce m_1 se odhaduje na základě dat (výnosů a dob do splatnosti) pro všechny vládní obligace EURO-bloku obchodované v daný den pomocí jádrového odhadu (viz SIMONOFF, 1996) a poté, s využitím hodnot již odhadnuté funkce $\hat{m}_1$, jsou odhadnuty individuální složky $\hat{m}_2^j$, v našem případě pro $J = 10$ států EMU (Lucembursko vládní obligace nemá).

Pro konstrukci testové statistiky zaokrouhlujeme pozorované doby do splatnosti t_i^j, $i = 1, \dots, n_j$, $j = 1, \dots, J$ na I hodnot t_i. Testová statistika je vhodně váženým součtem čtverců centrovaných částečných součtů

$$S_I = J \sum_{j=1}^{J} \sum_{k=1}^{I} (A_{jk} - \overline{A}_k)^2 (I\tau^2(j))^{-1},$$

kde $A_{jk} = \sum_{i=k+1}^{k+l} \hat{m}_2^j(t_i)$, $\overline{A}_k = \sum_{j=1}^{J} A_{jk}/J$ a $\hat{m}_2^j(t_i) = \hat{m}_2^j(t_{i-I})$ pro vhodně zvolené l a pro $I + 1 \leqq i \leqq I + l$; $\tau^2(j)$ jsou váhy pro stát j, které zohledňují variabilitu dat.

Hypotézu nerozlišitelných efektů jednotlivých států zamítáme pro velké hodnoty S_I. Pro odvození kritické hodnoty testové statistiky se osvědčil bootstrap. Jeho aplikace znamená, že počítáme hodnoty testové statistiky opakovaně pro mnoho souborů dat získaných pomocí prostého náhodného výběru z výchozí populace $\mathcal{Y} = \{\boldsymbol{y}_i,\ i = 1, \dots, I\}$ J-rozměrných vektorů o složkách $y_{ji} = \hat{m}_2^j(t_i) - \sum_{k=1}^{I} \hat{m}_2^j(t_k)/I$ a kritickou hodnotou je $(1 - \alpha)$-kvantil takto získané bootstrapové populace. Je-li hypotéza zamítnuta, můžeme testovat přítomnost posunutí některých individuálních výnosových křivek.

Metoda odhalila přetrvávající významné rozdíly v hodnocení jednotlivých států a naznačila jejich seskupení. Porovnání výsledků ze závěru roku 1998 a počátku roku 1999 svědčí pro zmenšování těchto rozdílů.

Literatura

[1] ABAFFY, J. et al. (1999): *A nonparametric model for analysis of the EURO yield curve.* Zasláno do J. Economic Dynamics and Control; viz také Technical Report 41, Universita Bergamo, Itálie.

[2] ANDERSEN, P. K., BORGAN, Ø. (1985): *Counting process models for life history data: a review.* Scand. J. Statist. *12*, 97–158.

[3] ANDERSEN, P. K. et al. (1993): *Statistical Models Based on Counting Processes.* Springer-Verlag, Berlin.

[4] BANERJEE, A. et al. (1993): *Co-integration, Error-Correction and the Econometric Analysis of Non-stationary Data.* Oxford University Press, Oxford.

[5] BOLLERSLEV, T. (1986): *Generalized autoregressive conditional heteroscedasticity.* J. of Econometrics *31*, 307–327.

[6] DARNELL, A. C. (1994): *A Dictionary of Econometrics.* Edwar Elgar, New York.

[7] DHRYMES, P. J. (1974): *Econometrics: Statistical Foundations and Applications.* Springer-Verlag, Berlin.

[8] DHRYMES, P. J. (1978): *Introductory Econometrics.* Springer-Verlag, Berlin.

[9] DHRYMES, P. J. (1994a): *Topics in Advanced Econometrics I — Probability Foundations.* Springer-Verlag, Berlin.

[10] DHRYMES, P. J. (1994b): *Topics in Advanced Econometrics II — Linear and Nonlinear Simultaneous Equations.* Springer-Verlag, Berlin.

[11] DICKEY, D. A., FULLER, W. A. (1979): *Distribution of the estimators for autoregressive time series with a unit root.* J. Amer. Stat. Assoc. *74*, 427–431.

[12] DURBIN, J., KOOPMAN, S. J. (2000): *Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives.* J. Roy. Stat. Soc. B. *62*, 3–56.

[13] EFRON, B. (1979): *Bootstrap methods: Another look at the jackknife.* Ann. Statist. *7*, 1–26.

[14] EFRON, B., TIBSHIRANI, R. J. (1993): *An Introduction to the Bootstrap.* Chapman & Hall, New York, London.

[15] ENGLE, R. F. (1982): *Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation.* Econometrica *50*, 987–1006.

[16] ENGLE, R. F. (1995): *ARCH. Selected Readings.* Oxford University Press, Oxford.

[17] ENGLE, R. F., GRANGER, C. W. J. (1987): *Cointegration and error correction: representation, estimation, and testing.* Econometrica *55*, 251–276.

[18] ENGLE, R. F., GRANGER, C. W. J., eds. (1991): *Long-Run Economic Relationships.* Oxford University Press, Oxford.

[19] FLEMING, T. R., HARRINGTON, D. P. (1991): *Counting Processes & Survival Analysis.* Wiley, New York.

[20] FOMBY, T. B., HILL, R. C., JOHNSON, S. R. (1984): *Advanced Econometric Methods.* Springer-Verlag, Berlin.

[21] GOURIÉROUX, CH. (1997): *ARCH Models and Financial Applications.* Springer-Verlag, Heidelberg–Berlin–New York.

[22] KALMAN, R. E. (1960): *A new approach to linear filtering and prediction problems.* J. of Basic Engineering *82*, 34–45.

[23] KENNEDY, P. (1998): *A Guide to Econometrics* (Fourth Edition). Blackwell Publishers, Cornwall.

[24] KMENTA, J. (1990): *Elements of Econometrics* (Second Edition). Macmillan, New York.

[25] KOENKER, R. (1982): *Robust methods in econometrics.* Econometric Reviews *1*, 213–255.

[26] KOKOSZKA, P., LEIPUS, R. (1998): *Covariance structure and change-point problem for non-negative ARCH processes.* In: *Prague Stochastics'98*, Vol. II, 321–324, JČMF Praha.

[27] SHAO, J., TU, D. (1995): *The Jackknife and Bootstrap.* Springer-Verlag, Heidelberg–Berlin–New York.

[28] SIMONOFF, J. S. (1996): *Smoothing Methods in Statistics.* Springer-Verlag, Berlin.

[29] VOŠVRDA, M., FILÁČEK, J., KAPIČKA, M. (1998): *The efficient market hypothesis testing on the Prague Stock Exchange.* Bull. of the Czech Econometric Society *7*, 55–67.