

KOINTEGRAČNÍ ANALÝZA NESTACIONÁRNÍCH VELIČIN

Jan Morávka

Třinecký inženýring, a.s., Třinec

Abstrakt

Cílem příspěvku je prezentovat nebezpečí tzv. *zdánlivé, nesmyslné, náhodné, klamné regrese* v případě *nestacionárních* (tzv. integrovaných) veličin. Nestacionární a integrované veličiny se vyskytují v technologii, buď jako výstupní veličiny z proporcionálních soustav s velkou setrvačností (nestacionární přechodové jevy a veličiny), nebo hlavně jako výstupní veličiny z integračních soustav (zásobníky, nádrže – integrované veličiny).

Zdánlivá regrese se vyskytuje v případě, že mezi nestacionárními veličinami neexistuje efekt tzv. *kointegrace*, což je skutečný dlouhodobý stabilní vztah (equilibrium) mezi veličinami. Takováto regrese může pak mít neblahý dopad na nesprávnou identifikaci, interpretaci a použití výsledků v technické a technologické praxi.

1 Úvod

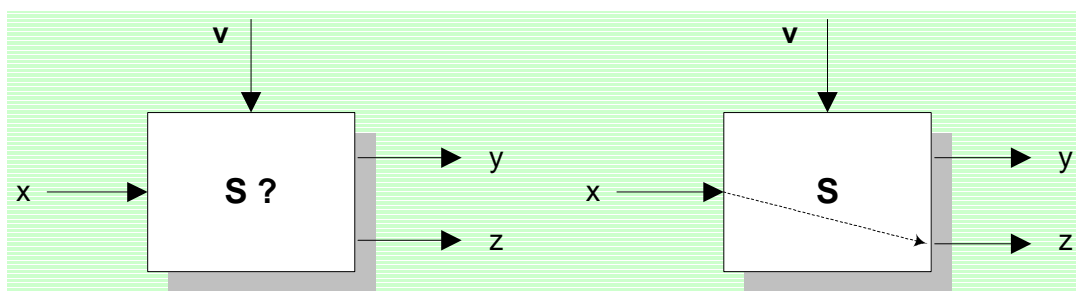
Pokud neexistuje mezi nestacionárními veličinami systému tzv. *kointegrace*, dochází k efektu tzv. *zdánlivé*, či *klamné* regrese. Její indikace, s dopadem na nesprávnou interpretaci závislostí, se vyskytuje u „expertního“, „okometrického“ odhadu a „rutinního“ použití klasické regresní analýzy.

Přístup a principy kointegrační analýzy jsou dokumentovány na nestacionárních vstupně-výstupních veličinách (časových řadách) modelovaného a simulovaného systému typu MIMO(1,2).

Pro analýzu *nestacionarity* (integrovanosti řádu jedna) veličin a jejich kointegrace byly vyvinuty speciální testy, které jsou aplikovány pouze v ekonometrických – a ne v běžných statistických – programech. V příspěvku jsou použity výsledky z freeware ekonometrických programů *Gretl* i *EasyReg* a z m-souborů programu MATLAB, obsažených jak v public domain balících *Econometrics Toolbox* a *Time Series Statistics*, tak vytvořených autorem tohoto příspěvku.

2 Model a data

Jako model byl zvolen systém S typu MIMO(1,2) = SIMO(2), tj. systém s jedním vstupem (x), dvěma výstupy (y , z) a vektorem vnějších veličin (v) – obr.1:

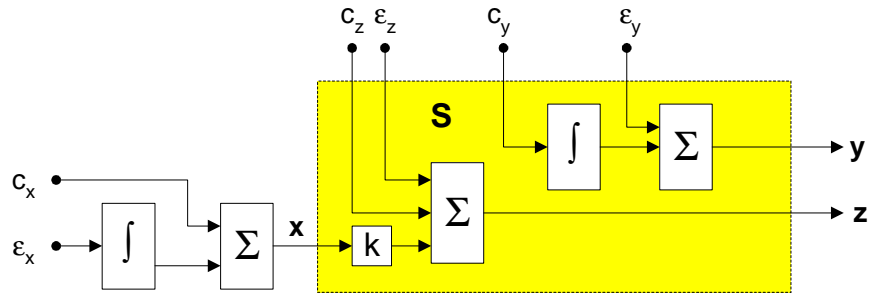


Obr. 1. Vnější blokové schéma simulovaného systému SIMO(2) – „černá“ a „šedá“ skříňka

V levé části obr.1 je systém znázorněn jako tzv. *černá* skříňka (bez znalosti vnitřní struktury a vazeb mezi signály) a v pravé části jako tzv. *šedá* skříňka s vyznačením vazby mezi veličinami x a z . Mezi veličinami x a y žádná vazba neexistuje.

Vstupní a výstupní veličiny mají charakter *nestacionárních* procesů typu I(1), tj. tzv. *integrovaných* veličin 1. řádu. Vnější veličiny mají charakter základní *stacionární* veličiny typu I(0), což je tzv. gaussovský bílý šum a konstantní signál [ARLT 1999].

Podrobné vnitřní blokové schéma systému ve tvaru tzv. *bílé* skřínky (se spojitě v čase pracujícími bloky) je znázorněné na *obr.2* (kde symbol \int označuje integrátor – integrační člen a symbol Σ sumátor – sumační, sčítací člen).



Obr. 2. Podrobné vnitřní schéma simulovaného systému SIMO(2)

Rovnice, veličiny a konstanty modelovaného a simulovaného systému jsou v obecnějším tvaru následující:

$$x \sim ARIMA(0,1,0; \sigma_x) + c_x = I(1; \sigma_x) + c_x = RW(\sigma_x) + c_x, \quad (1a)$$

$$y \sim ARIMA(0,1,0; \sigma_y) + c_y = I(1; \sigma_y) + c_y = RW(\sigma_y) + c_y, \quad (1b)$$

$$z = k \cdot x + \varepsilon_z + c_z \sim I(1; k \cdot \sigma_x) + \varepsilon_z + c_z = RW(k \cdot \sigma_x) + \varepsilon_z + c_z, \quad (1c)$$

- kde je
- c_i - konstantní signál, $i \in \{x, y, z\}$,
 - ε_i - gaussovský bílý šum $\sim WN(0, \sigma_i)$,
 - σ_i - směrodatná odchylka gaussovského bílého šumu,
 - ARIMA(p,d,q) - nestacionární smíšený integrovaný proces řádu $\{p, d, q\}$,
 - I(1) - integrovaný proces 1. řádu = ARIMA(p,1,q),
 - RW - „náhodná procházka“, Random Walk = ARIMA(0,1,0),
 - k - koeficient přenosu (zesílení).

Je třeba zdůraznit, že všechny tři bílé šумы jsou vzájemně nekorelované (vzájemně nezávislé), tj. platí $cov(\varepsilon_x, \varepsilon_y) = cov(\varepsilon_x, \varepsilon_z) = cov(\varepsilon_y, \varepsilon_z) = 0$.

Diskrétní rovnice, popisující generování veličin (časových řad) simulovaného systému mají výpočtový tvar:

$$sx_t = sx_{t-1} + \varepsilon_{xt}; \quad x_t = sx_t + c_x, \quad (2a)$$

$$sy_t = sy_{t-1} + \varepsilon_{yt}; \quad y_t = sy_t + c_y, \quad (2b)$$

$$z_t = k \cdot x_t + \varepsilon_{zt} + c_z, \quad (2c)$$

- kde je
- t - diskrétní čas, $t = 1, 2, \dots, n$,
 - n - počet hodnot, délka časové řady, $n \in \mathbb{N}$,
 - si_t - sumační pomocná proměnná, $si_{t=0} = si_0 = 0$,
 - i_t - hodnoty časové řady v čase t.

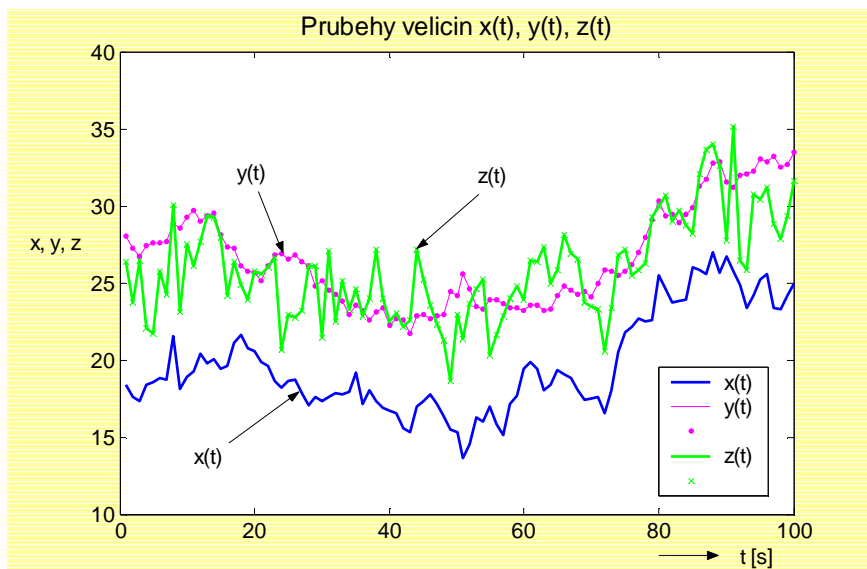
Konkrétní hodnoty konstant, koeficientů a parametrů jsou u vedeny v *tab.1*:

Tab. 1. PARAMETRY SIMULACE

Parametr	Hodnota	Poznámka
n	100	počet hodnot
c_x	20	
σ_x	1.0	
c_y	25	
σ_y	0.5	
c_z	10	
σ_z	0.4	
k	0.8	zesílení

Hodnoty veličin $x(t)$ a $y(t)$ systému S byly vygenerovány jako diskrétní časové řady typu ARIMA(0,1,0) a veličina $z(t)$ jako lineární transformace veličiny $x(t)$ (s pomyslnou periodou vzorkování $T = 1$ s) v programu MATLAB, kde byly také uloženy do textového souboru *Coint.dat*. Tento „univerzální“ tvar souboru byl vhodný pro další použití, a to jak v tabulkovém procesoru *Excel*, tak i v ekonometrických freeware (GPL) programech *Gretl* [COTTRELL 2005], [KUFEL 2004] a *EasyReg* [BIERENS 2005].

Na obr.3 je znázorněný časový průběh vstupně-výstupních veličin systému S.



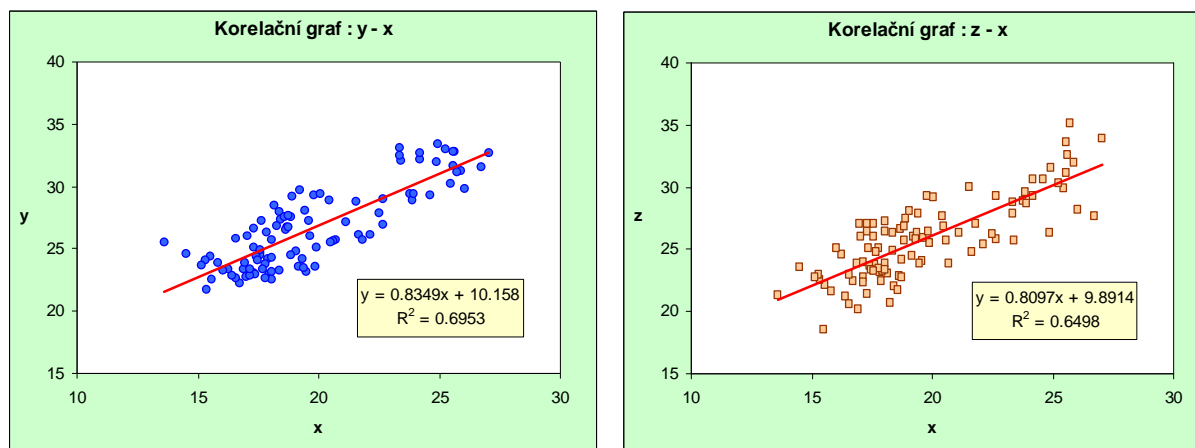
Obr. 3. Časový průběh veličin systému S

Níže uvedená analýza (veličin) systému je seřazena podle nesprávnosti (až nakonec po správnosti) přístupů k řešení: počínaje „okometrickým“ přístupem, přes povrchní, klasickou a moderní lineární regresní analýzu, jako i empirickou indikaci zdánlivé regrese, až k samotné kointegrační analýze, zahrnující testování jednotkového kořene veličin a kointegraci dvojic veličin.

3 „Okometrická“ analýza

Pokud se pouze „expertně a okometricky“ (a bohužel povrchně) podíváme na časový průběh veličin systému na obr.3, pak lze konstatovat, že obě výstupní veličiny $y(t)$ a $z(t)$ mají podobný průběh (trend) jako vstupní veličina $x(t)$, a tedy mezi nimi zřejmě existuje závislost.

Na potvrzení této domněnky si „expert“ může ještě v tabulkovém procesoru *Excel* sestavit bodové (korelační) X-Y grafy a proložit jimi regresní přímky s uvedením regresních rovnic a koeficientů determinace R^2 – obr.4:



Obr. 4. Korelační grafy veličin systému s regresními přímkami

Výsledky ho „přesvědčí“, že skutečně mezi vstupně-výstupními veličinami systému, tj. mezi dvojicemi $y(t)$ a $x(t)$, jako i mezi $z(t)$ a $x(t)$ existuje korelace a lineární regresní závislost.

Koeficienty determinace R^2 (podíl vysvětleného rozptylu výstupní veličiny modelem) a případně tzv. regresní rabaty (což jsou 100-násobky koeficientů determinace pro jejich názornější vyjádření v procentech) mají hodnoty asi 70 % a 65 %, které jsou pro technickou a technologickou praxi „dostatečně velké“. Z jejich hodnot by navíc vyplývalo, že závislost mezi veličinami $y(t)$ a $x(t)$ je těsnější než závislost mezi $z(t)$ a $x(t)$. Veličina $y(t)$ má navíc vůči veličině $z(t)$ hladší (méně „zašuměný“) průběh, co by „expertovi“ potvrdilo její přednost při případném rozhodnutí o výběru jedné ze jmenovaných výstupních veličin.

4 Lineární regresní analýza

Ještě zkušenější „expert“ na statistiku provede opět „pro jistotu“ ještě lineární regresi v tabulkovém procesoru *Excel* nebo v nějakém běžném statistickém programu (např. v programu *Statgraphics*, či *Statistica*), či v kvalitním statistickém programu (např. *QC Expert*) s bohatou a úplnou regresní diagnostikou.

Lineární regresní model je přitom uvažován v jednoduchém základním tvaru (ve variantách pro obě výstupní, čili vysvětlované veličiny – regresandy, přičemž z důvodu zjednodušení zápisu nejsou uváděny časové indexy u veličin, tj. $y(t) = y_t = y$, $z(t) = z_t = z$, $x(t) = x_t = x$):

$$y = a_{yx} + b_{yx} \cdot x + e_{yx}, \quad (3a)$$

$$z = a_{zx} + b_{zx} \cdot x + e_{zx}, \quad (3b)$$

kde je

- a - absolutní člen,
- b - regresní koeficient,
- e - reziduum regresního modelu.

Jako *metoda* výpočtu regresních koeficientů je uvažována nejužívanější metoda nejmenších čtverců (MNC) a hladina statistické významnosti testů $\alpha = 0,05$. Při podkročení dosažené (empirické) hladiny významnosti pod stanovené α jsou dále prezentované hodnoty koeficientů a testovacích statistik uvedeny v závorce.

4.1 Povrchní regresní analýza

Pokud uvažujeme obvyklé (a přitom bohužel pouze základní a nedostatečné) výstupy lineární regrese z tabulkového procesoru *Excel* nebo z běžně užívaných statistických programů, dostaneme výsledky uvedené v *tab.2* (u použití *Excelu* při volbě *Nástroje/Analýza dat/Regrese*, výsledky jsou přirozeně totožné s hodnotami na předchozím *obr.4*):

Tab. 2. VÝSLEDKY POVRCHNÍ REGRESNÍ ANALÝZY

Objekt	Parametr / vlastnost	Koeficient / test	Hodnota	Hodnocení
Model	regresní koeficienty	a_{yx} / a_{zx}	10.158 / 9.891	významné / významné
		b_{yx} / b_{zx}	0.835 / 0.810	významné / významné
	významnost	Fisher-Snedecorův	223.7 / 181.8	významné / významné
		R^2 [%]	69.5 / 65.0	vyšší / vyšší

Z tabulky je zřejmé, že regresní model i všechny regresní koeficienty jsou *pro obě dvojice veličin statisticky významné*.

Příkladem možné - a přitom *nesprávné a nesmyslné interpretace* výsledků (s případným neblahým dopadem do technologické praxe) - může být následující tvrzení:

- obě závislosti jsou statisticky významné a mají prakticky stejné hodnoty regresních koeficientů.

Jelikož závislost $y(t) - x(t)$ je těsnější, „přesnější“ (má vyšší koeficient determinace asi o 5 %), stačí dále uvažovat pouze tuto závislost, která tak zastupuje „duplicitní, nadbytečnou, nepotřebnou“ závislost $z(t) - x(t)$.

Tento závěr je však naprosto chybný, protože u fyzicky neexistující vazby mezi veličinami $y-x$ jde o efekt tzv. klamně, zdánlivé, a v daném případě tedy i *nesmyslné regrese*.

4.2 Moderní regresní analýza

Ještě zkušenější „expert“ na statistiku uskuteční lineární regresní analýzu v kvalitním statistickém programu, který obsahuje *regresní diagnostiku* včetně *analýzy reziduí*, s komplexním testováním tzv. regresního tripletu – data, model, metoda [MELOUN & MILITKÝ 1994, 2004]. Výsledky ze statistického systému *QC Expert* jsou uvedeny v *tab.3a* a *tab.3b*.

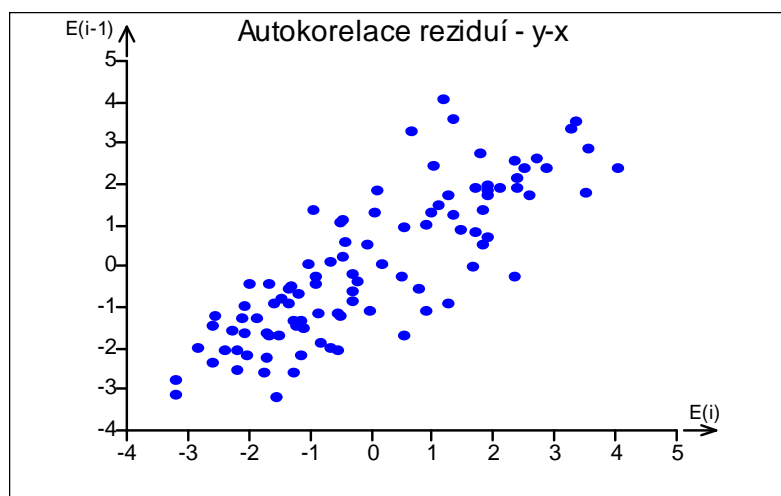
Tab. 3a. VÝSLEDKY MODERNÍ REGRESNÍ ANALÝZY PRO MODEL Y-X

Objekt	Parametr / vlastnost	Koeficient / test	Hodnota	Hodnocení
Model	regresní koeficienty	a_{yx}	10.158	významný
		b_{yx}	0.835	významný
	významnost	Fisher-Snedecorův	223.7	významný
		R^2	0.695	69.5 %, vyšší
korektnost	Scottovo kritérium	0.186	korektní	
Rezidua	homoskedasticita	Cook – Weisbergův	0.053	homoskedasticita
	normalita	Jarque – Berraův	5.091	normalita
	náhodnost	Waldův	179.2	významná autokorelace!
		Durbin – Watsonův	0.341	pozitivní autokorelace!
	trend	znaménkový	7.095	trend!

Pozn.: Kritické hodnoty Durbinova – Watsonova testu jsou pro $\alpha = 0.05$, $n = 100$, $k = 1$ (počet regresorů bez absolutního členu) $dL = 1.65$, $dU = 1.69$ – viz např. program *Gretl* [COTTRELL 2005]. Rozsáhlé a přesné kritické hodnoty pro $\alpha = 0.05$, $n = 6 \div 2000$, $K = 2 \div 21$ (počet regresorů s absolutním členem, $K = k + 1$) je možné nalézt v [CUMMINS 2005].

Z *tab.3a* je zřejmé, že regresní model $y-x$ a oba regresní koeficienty jsou sice *statisticky významné*, avšak regresní model *není statisticky korektní* – rezidua vykazují významnou pozitivní autokorelaci (*obr.5*) a trend, co signalizuje:

- buď nesprávný (v krajním případě nesmyslný) model,
- nebo opomenutí (nezahrnutí) v modelu dalšího regresandu (vysvětlující proměnné).



Obr. 5. Graf autokorelace reziduí modelu $y-x$

Znova je však nutné konstatovat, že i tento závěr *není správný* – model $y-x$ a jeho koeficienty

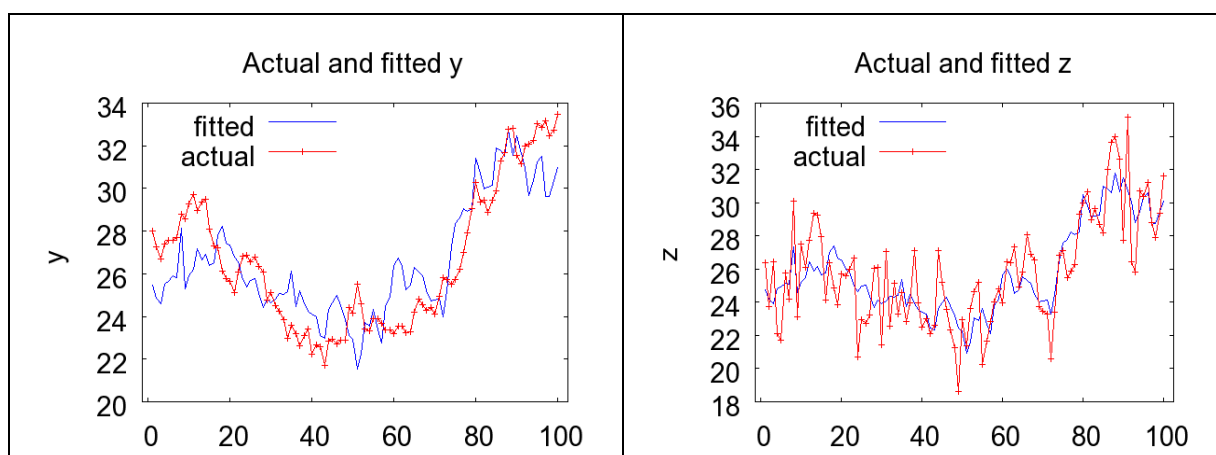
nemohou být statisticky významné, protože model de facto neexistuje (mezi veličinami $y(t)$ a $x(t)$ není žádné fyzikálně-matematické propojení, žádná vazba). Je tedy možné vyslovit vážné podezření na výskyt efektu *zdánlivé regrese*.

Tab. 3b. VÝSLEDKY MODERNÍ REGRESNÍ ANALÝZY PRO MODEL Z-X

Objekt	Parametr / vlastnost	Koeficient / test	Hodnota	Hodnocení
Model	regresní koeficienty	a_{zx}	9.891	významný
		b_{zx}	0.810	významný
	významnost	Fisher-Snedecorův	181.8	významný
		R^2	0.645	64.5 %, vyšší
	korektnost	Scottovo kritérium	0.185	korektní
Rezidua	homoskedasticita	Cook – Weisbergův	0.928	homoskedasticita
	normalita	Jarque – Berraův	1.694	normalita
	náhodnost	Waldův	0.766	není autokorelace
		Durbin – Watsonův	1.810	není autokorelace
	trend	znaménkový	1.504	není trend

Z tab.3b vyplývá, že regresní model z-x i oba regresní koeficienty jsou *statisticky významné* a regresní model je *statisticky korektní*. Tento závěr je i fakticky správný.

Pro srovnání je grafické zobrazení odhadnutých (predikovaných, fitovaných) a původních hodnot veličin $y(t)$ a $z(t)$ obou modelů viditelné na obr.6 (Gretl):



Obr. 6. Predikční grafy výstupních veličin regresních modelů y-x a z-x

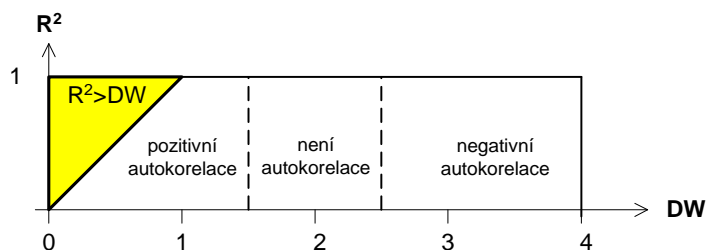
4.3 Signalizace zdánlivé regrese

V literatuře [ARLT 1997], [GAJDA 2004] je uveden empirický poznatek o souvislosti mezi vysokými hodnotami F-statistik modelů, jako i t-statistik regresních koeficientů a nízkými hodnotami Durbinovy-Watsonovy (DW) statistiky reziduí u *zdánlivé regrese*. Je to natolik charakteristická vlastnost *zdánlivé regrese*, že *Granger a Newbold (1974)* navrhli, aby splnění nerovnosti:

$$\boxed{R^2 > DW}, \tag{4}$$

tj. když koeficient (obecně index) determinace je větší než DW statistika, bylo určitým indikátorem nebezpečí existence *zdánlivé regrese* – viz i obr.7:

Obr. 7. Kritická oblast *zdánlivé regrese*



Na základě údajů uvedených v *tab.3a,b* je zřejmé, že:

- nerovnost (4) je splněna u regrese $y-x$, jelikož $R^2 > DW$, tj. $0.695 > 0.341$, což signalizuje možnost existence zdánlivé regrese. Tento závěr je navíc umocněn výskytem významné pozitivní autokorelace a trendu reziduí. Z uvedených skutečností tedy vyplývá, že regresní model $y-x$ je nesprávný a dokonce nesmyslný, protože neexistující, což odpovídá vnitřní struktuře modelu S,
- u regrese $z-x$ platí opačná relace: $R^2 < DW$, tj. $0.645 < 1.810$, která nesignalizuje vznik zdánlivé regrese, což je opět umocněno skutečností náhodnosti (nekorelovanosti) reziduí. Tento regresní model se jeví jako správný, statisticky korektní a existující.

4.4 Závěr k lineární regresní analýze

Pro rozpoznání a signalizaci nebezpečí existence zdánlivé regrese je nutné:

- použít SW obsahující výpočet Durbinovy-Watsonovy statistiky DW pro stanovení autokorelace reziduí 1. řádu (nebo tuto statistiku vypočítat v tabulkovém procesoru Excel, či v SW MATLAB podle vztahu uvedeného např. v [MELOUN & MILITKÝ 2004], [ARLT 1999] nebo [MARČEK & MARČEK 2001]),
- pokud je prokázána statisticky významná pozitivní autokorelace reziduí, tj. hodnota DW je kladná, nízká a pod kritickou mezí, pak je možné předpokládat *nesprávný* regresní model. Pokud je navíc splněno i Grangerovo-Newboldovo pravidlo (nerovnost), pak může být takovýto regresní model i *zdánlivý*, *klamný* a dokonce *nesmyslný*.

Pro korektní rozlišení zdánlivé regrese je však nutné testovat tzv. *integrovanost* (stochastickou *nestacionaritu*) časových řad a následně jejich *kointegraci*.

5 Kointegrační analýza

Kointegrační analýza je moderní disciplína vyvinuta na přelomu 80. a 90. let 20. století pro potřeby analýzy *nestacionárních* veličin (časových řad) v ekonometrických modelech [ARLT 1997, 1999], [MARČEK & MARČEK 2001]. Dá se říci, že výrazně změnila charakter ekonometrické analýzy časových řad.

Kointegrační analýza může být uskutečněna více způsoby - buď korektně pomocí *numerických* testů a *kointegrační regresní rovnice* anebo přibližně, ale zato názorně, pomocí *grafického* znázornění.

V dalším textu bude ukázán pouze základní způsob pomocí *numerické kointegrační analýzy*, kde její postup pozůstává ze dvou následujících kroků:

1. Testování integrovanosti veličin – test $I(1)$ = test tzv. jednotkových kořenů.
2. Testování výskytu kointegrace dvou veličin – test $CI(1,1)$.

Prakticky to znamená, že kointegraci dvou (obecně m) veličin má význam testovat jen tehdy, pokud jsou obě veličiny *nestacionární* a tzv. integrované alespoň řádu 1. Tuto skutečnost lze zjistit právě pomocí testů jednotkových kořenů.

5.1 Testování jednotkových kořenů

Testování tzv. jednotkových kořenů slouží ke stanovení typu náhodné veličiny, tj. zda veličina je *nestacionárním* procesem typu $I(1)$, tzn. integrovaným procesem 1. řádu. Časová řada je typu $I(1)$, když jeho diference je obecná *stacionární* časová řada typu $I(0) = ARMA(p,q)$, ve speciálním a nejjednodušším případě je to tzv. bílý šum $WN = ARMA(0,0) = AR(0) = MA(0)$.

Pokud uvažujeme jednoduchý *stacionární* $AR(1)$ proces typu $I(0)$:

$$\boxed{y_t = \rho \cdot y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_\varepsilon^2)}, \quad (5)$$

kde je ρ - koeficient autokorelace 1. řádu, autoregresní koeficient 1. řádu,

ε - reziduum typu gaussovský bílý šum s rozptylem σ_ε^2 ,

pak se tento proces stane *nestacionárním* typu I(1), když parametr $\rho = 1$ (tzv. *jednotkový kořen*) a takovýto proces má v literatuře označení „náhodná procházka“ (RW - Random Walk). RW obsahuje tzv. stochastický (nedeterministický) trend, jelikož jde o nestacionární proces v *rozptylu*, přičemž rozptyl roste přímo úměrně s časem (délkou) časové řady, tj. $\text{var}(y_t) = t \cdot \sigma_\varepsilon^2$ [ARLT 1999], [GAJDA 2004].

Na testování (nulové) hypotézy $H_0: \rho = \rho_0 = 1$ (a na hypotézu alternativní – testování stacionarity) existuje několik parametrických a neparametrických testů:

- ❑ mezi *parametrické* testy patří testy - základní Dickey-Fullerovy (má 2 varianty: DF(z = f(ρ)) a DF(t) statistiky), rozšířené ADF (Augmented Dickey-Fuller) a Said-Dickeyovy (SD),
- ❑ mezi testy *neparametrické* lze zařadit test Phillipsův, testy Phillips-Perronovy, Newey-Westovy, Bierensovy, Bierens-Guovy a alternativní KPSS (Kwiatkowski – Phillips – Schmidt – Shin).

Blíže jsou uvedené testy popsány v literatuře [ARLT 1999], [ARLT & ARLTOVÁ 2003], [MARČEK & MARČEK 2001], [BIERENS 2005], [COTTRELL 2005], [NEUBAUER 2005].

Pro analyzované časové řady byl aplikován nejčastěji používaný a v programech implementovaný ADF test.

ADF test

Tento test je poměrně obecný, protože předpokládá autokorelační strukturu reziduí ve formě procesu AR(p). Nulová hypotéza je testována pro tři typy (varianty) regresních modelů obecného tvaru:

$$y_t = \mu + \rho \cdot y_{t-1} + \delta \cdot t + \sum_{i=1}^p \gamma_i \Delta y_{t-i} + \varepsilon_t. \quad (6)$$

Jde o varianty: model bez konstanty μ a deterministického trendu $\delta \cdot t$ (ADF-1), model s konstantou (ADF-2) a úplný model (ADF-3). Pro každý z modelů jsou v tabulkách – uvedených např. v [ARLT 1999], [MARČEK & MARČEK 2001] – stanoveny jiné kritické hodnoty.

Volba řádu modelu AR(p), tj. p , se dá stanovit podle autorů Newey-West [BIERENS 2005] jako dolní celá část výrazu $[c \cdot n^r]$, kde je doporučena volba $c = 5$, $r = 0.25$.

Podle charakteru průběhu veličin x, y, z volíme model ADF-2 (s nenulovou konstantou a bez deterministického trendu), přičemž řád modelu AR je $p = [15.811] = 15$.

Kritické hodnoty jsou pro $\alpha = 0,05 = 5\%$ a $n = 100$ následující: DF-test má t-krit = **-2.89** a ADF-2 test má (limitní) t-krit = **-2.86** [ARLT 1999], [MARČEK & MARČEK 2001], podle m-souborů *adf.m* a *zcrit.m* z Econometrics Toolbox (ET) – viz literatura [LESAGE 1999], je příslušná kritická hodnota t-krit = **-2.92**.

Pokud $t\text{-stat} > t\text{-krit}$, pak nelze zamítnout nulovou hypotézu o integrovanosti veličin řádu 1.

Pro výše uvedené parametry a veličiny dostaneme z m-funkcí *adf*, *cadf*, *zcrit* a *rzcrit* z balíku ET určité výsledky.

Obdobně lze použít i funkci *adfreg.m* a *dfcrit.m* z balíku TS_STAT [KANZLER 1998].

Výsledky z obou toolboxů se však (dost) liší od prakticky stejných výsledků z ekonometrických programů *EasyReg* a *Gretl*, které navíc obsahují také p-value t-statistik (vycházející z aproximací [MACKINNON 1996]), což je jednodušší a zřejmější pro posouzení platnosti nulové hypotézy – viz *tab.4*:

Tab. 4. VÝSLEDKY TESTOVÁNÍ JEDNOTKOVÝCH KOŘENŮ VELIČIN

Veličina	ET	TS_STAT		Gretl, EasyReg		Hodnocení
	t-stat <i>adf.m</i>	t-stat <i>adfreg.m</i>	t-stat <i>unitroot.m: ADF/DF</i>	t-stat	p-value	
x	-0.68	-1.36	-1.24 / -1.24	-1.22	0.66	x ~ I(1)
y	-1.03	-1.02	0.10 / 0.10 ??	-0.92	0.78	y ~ I(1)
z	-4.29 ??	-0.15	-2.70 / -4.41 ??	-0.13	0.94	z ~ I(1)

Jak je z *tab.4* zřejmé, funkce *adf.m* z ET a funkce *unitroot.m* z TS_STAT poskytují nespolehlivé a v podstatě nesprávné výsledky.

Naopak, podle shodných a spolehlivých výsledků programů *Gretl* a *EasyReg*, jako i přibližně stejných výsledků funkce *adfreg.m*, nelze zamítnout nulovou hypotézu u všech veličin a tedy jsou integrované 1. řádu, tj. *nestacionární* v rozptylu (obsahují tzv. *stochastický trend*). Pro takovéto veličiny má smysl provést následnou *kointegrační* analýzu.

5.2 Testování výskytu kointegrace

Testování výskytu kointegrace bude uskutečněno po dvojicích vstupně-výstupních nestacionárních veličin, které jsou typu I(1).

Existenci kointegrace dvou veličin 1. řádu (ve smyslu dlouhodobého vztahu, tzv. *ekvilibria*) je označována jako CI(1,1) – viz např. [ARLT 1999], [MARČEK & MARČEK 2001], obecně pak CI(d,q).

Pro testování kointegrace existuje vícero testů: CRDW – Durbinův-Watsonův, CRDF – Dickeyův-Fullerův (se dvěma variantami z = f(p), t), CRADF – Augmented DF, Phillipsův, Johansenův, Engle-Grangerův a Bierensův. Uvedené testy jsou podrobně popsány v literatuře [MARČEK & MARČEK 2001], [ARLT 1999], [BIERENS 2005], [COTTRELL 2005], [NEUBAUER 2005].

Pro analyzovaná simulační data použijeme pro jednoduchost pouze nejjednodušší a nejnázornější test CRDW, i když málo používaný – z dostupné literatury byl uveden pouze v [MARČEK & MARČEK 2001] a [BENKOVSKIS 2004].

CRDW test

Při použití *CRDW* (*C*ointegrating *R*egression *D*urbin-*W*atson) testu [MARČEK & MARČEK 2001] nejprve vypočteme parametry statické lineární regrese (*bez absolutního členu*, je však možné i s absolutním členem) a stanovíme klasickou statistiku DW reziduí, která je výpočetně shodná ze statistikou CRDW – viz *tab.5a,b* (výstupy jsou z programu *Gretl*, v závorce jsou uvedeny hodnoty DW statistiky pro regresi s absolutním členem):

$$y = b_{yx} \cdot x + e_{yx}, \quad (7a)$$

$$z = b_{zx} \cdot x + e_{zx}. \quad (7b)$$

Tab. 5a. VÝSLEDKY STATICKÉ REGRESE PRO CRDW TEST U DVOJICE VELIČIN Y-X

Objekt	Parametr / vlastnost	Koeficient / test	Hodnota	Hodnocení
Model	regresní koeficient	b_{yx}	1.338	významné
	významnost	Fisher-Snedecorův	12075	významné
		R^2 [%]	69.5	vyšší
Rezidua	náhodnost	Durbinův-Watsonův DW = CRDW	0.377 (0.341)	<i>není kointegrace</i>

Tab. 5b. VÝSLEDKY STATICKÉ REGRESE PRO CRDW TEST U DVOJICE VELIČIN Z-X

Objekt	Parametr / vlastnost	Koeficient / test	Hodnota	Hodnocení
Model	regresní koeficient	b_{zx}	1.299	významné
	významnost	Fisher-Snedecorův	10737	významné
		R^2 [%]	65.0	vyšší
Rezidua	náhodnost	Durbinův-Watsonův DW = CRDW	1.108 (1.810)	kointegrace

Samotné testování kointegrace pomocí CRDW testu vychází z hypotéz: nulová hypotéza H_0 : veličiny *nejsou* kointegrované, rezidua mají charakter RW typu I(1) a alternativní hypotéza má tvar H_1 : veličiny jsou kointegrované, rezidua mají charakter *pozitivně* autokorelovaného procesu AR(1), který je typu I(0).

Pokud hodnota statistiky $\boxed{CRDW \leq CRDW\text{-krit}(m, \alpha, n)}$, pak nelze zamítnout nulovou hypotézu o neexistence kointegrace veličin. Parametry kritických hodnot přitom jsou: m – počet proměnných v kointegraci, α – hladina významnosti, n – počet hodnot proměnných.

Kritické hodnoty CRDW statistiky pro $\alpha = 0.05$, $m \in \{2, 3, 4, 5\}$, $n \in \{50, 100, 200\}$ jsou uvedeny v [MARČEK & MARČEK 2001]. Simulačně byly kritické hodnoty pro parametry: $\alpha = \{0.01, 0.05, 0.10\}$, $m \in \{2, 3, 4, 5\}$, $n \in (20, 200)$ stanoveny pomocí metody Monte-Carlo – viz [BENKOVSKIS 2004].

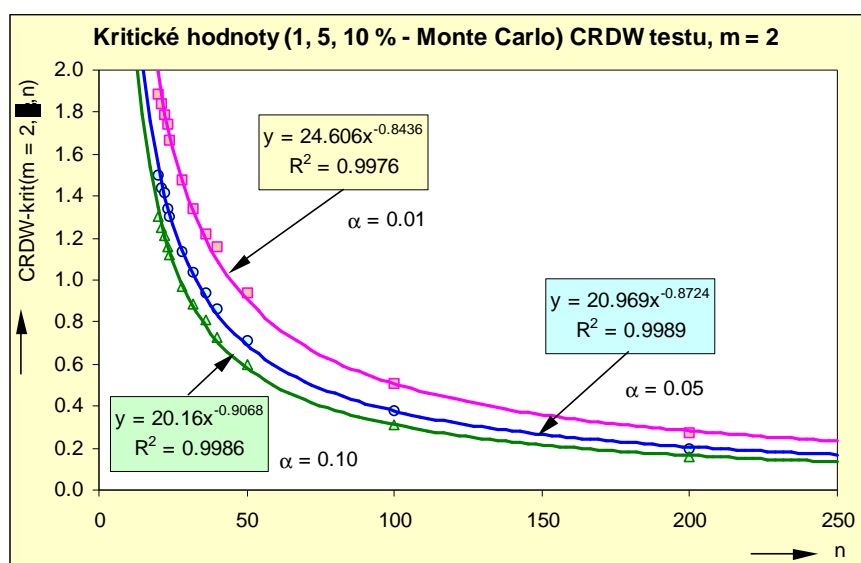
Na obr.8 je vidět průběhy a jednoduchá aproximace (pomocí mocninné funkce (8a)) kritických hodnot statistiky CRDW, vycházejících z výše uvedených simulačních výpočtů, pro hladiny významnosti $\alpha = 0.01, 0.05, 0.10$, $m = 2$, $n \geq 20$. Další vhodnou aproximací je reciproční lineární funkce (8b). Parametry obou aproximací jsou uvedeny v tab.6:

$$CRDW_{krit}(m = 2, \alpha, n) \approx a \cdot n^b, \quad (8a)$$

$$CRDW_{krit}(m = 2, \alpha, n) \approx \frac{1}{a \cdot n + b}. \quad (8b)$$

Tab. 6. PARAMETRY APROXIMACÍ KRITICKÝCH HODNOT CRDW STATISTIKY PRO $m = 2$, $n \geq 20$

Aproximace → ↓ α	CRDW-krit = $a \cdot n^b$		CRDW-krit = $1/(a \cdot n + b)$	
	a	b	a	b
0.01	24.606	-0.8436	0.0178	0.1696
0.05	20.969	-0.8724	0.0248	0.1704
0.10	20.160	-0.9068	0.0301	0.1689



Obr. 8. Průběhy a aproximace kritických hodnot CRDW testu pro $m = 2$

Pro $\alpha = 0.05$, $m = 2$, $n = 100$ je hodnota CRDW-krit = **0.386** [MARČEK & MARČEK 2001]. Z výsledků testování v *tab.5a,b* je zřejmé, že:

- veličiny $y-x$ nejsou kointegrované, protože $CRDW < CRDW\text{-krit}$, tj. $0.377 / 0.341 < 0.39$ a nulovou hypotézu o neexistenci kointegrace veličin nelze zamítnout (přítom lepší rozlišení poskytují hodnoty $CRDW = DW$ statistiky pro regresní model s *absolutním členem*),
- veličiny $z-x$ kointegrované *jsou*, protože $CRDW > CRDW\text{-krit}$, tj. $1.108 / 1.810 > 0.39$ a nulovou hypotézu o neexistenci kointegrace veličin tedy lze zamítnout (přítom i v tomto případě lepší rozlišení poskytují hodnoty $CRDW = DW$ statistiky pro regresní model s *absolutním členem*).

Hodnocení:

- CRDW test správně vyhodnotil existenci kointegrace mezi veličinami $z-x$ a současně její neexistenci u dvojice veličin $y-x$, co znamená že regrese je zde pouze zdánlivá a nesmyslná,
- pro lepší rozlišení dosažených a kritických hodnot CRDW testu je vhodnější užít lineární regresi s *absolutním členem*,
- CRDW test je výhodný, protože používá běžnou DW statistiku, je jednoduchý i názorný a pro stanovení jeho kritických hodnot lze použít jednoduché monotónní 2-parametrové aproximace (pomocí mocninné a reciproční lineární funkce). Podmínkou je však dodržení rozsahu hodnot alespoň $n \geq 20$.

Další testy kointegrace

Pro analyzovaná data byly použity ještě:

- další *numerické* testy: CRDF, Johansenův, Engle-Grangerův a Bierensův [ARLT 1999], [MARČEK & MARČEK 2001], [COTTRELL 2005], [BIERENS 2005], [NEUBAUER 2005],
- tzv. *kointegrační regresní rovnice* ve tvaru modelů ADL (Autoregressive Distributed Lag) – viz např. [LUCCHETTI 2005], [ARLT 1999], [MARČEK & MARČEK 2001] a ECM (Error Correction Model) – viz [ARLT 1999], [MARČEK & MARČEK 2001],
- *grafické* posouzení průběhu reziduí a jejich autokorelační (ACF) i parciální autokorelační funkce (PACF).

Principy a výsledky uvedených testů už nejsou v příspěvku uvedeny s ohledem na jeho čitelnost a přehlednost.

Nicméně souhrnně lze konstatovat, že testy CRDF a Johansenův (tento však pouze v případě uvažování VAR(1) modelu s absolutním členem), kointegrační regresní rovnice (modely ADL i ECM) a grafické posouzení charakteru reziduí správně detekovaly kointegraci dvojic veličin.

Na druhé straně však testy Engle-Grangera, Bierensa a také Johansena (tento při uvažování modelu bez absolutního členu) poskytly pouze „napůl“ správné (a tím nespolehlivé) výsledky.

Použití m-funkcí *cadf.m* (test CRADF) a *rztcrit.m* (příslušné kritické hodnoty testu) z balíku Econometrics Toolbox [LESAGE 1999] poskytlo obdobně nespolehlivé výsledky. Balík m-funkcí, vytvořených autorem [NEUBAUER 2005], se naopak osvědčil a výsledky funkcí byly totožné s výsledky z ekonometrických programů *Gretl* a *EasyReg*.

Autorem tohoto příspěvku byla vytvořena m-funkce *CRDWtest.m*, umožňující uskutečňovat zjednodušenou (přibližnou, orientační) kointegrační analýzu. Funkce počítá a používá DW statistiku, koeficient determinace, empirickou Grangerovu-Newboldovu nerovnost a výše uvedené aproximační kritické hodnoty CRDW statistiky.

6 Toolboxy pro ekonometrii a analýzu časových řad

V *tab.7* je uveden přehled toolboxů, vhodných pro analýzu časových řad s ohledem na její použití v ekonometrii. Jsou zde definovány potřebné funkce a testy pro tyto úlohy. Dostupné balíky jsou rozčleněny na skupiny podle dostupnosti (či přístupnosti, tj. ceny).

Tab.7. POROVNÁNÍ MOŽNOSTÍ TOOLBOXŮ PRO EKONOMETRII

Tbx → Statistika, test ↓	Komerční		Third-party products			Osobní	
	Stat	GARCH	ET	TS_STAT	TSA	NJ	
Autokorelace	DW	dwtest	-	-	dwatson	-	-
	Dh	-	-	-	durbinh	-	-
	ACF	-	autocorr	-	-	acorf	-
	PACF	-	parcorr	-	-	pacf	-
Jednotkové kořeny	DF	-	dfARTest, dfARDTest, dfTSTest	adf, ztcrit	dfcrit	-	unitrootx, x = 1-4
	ADF	-			adfreg, unitroot	-	-
	PP	-	ppARTest, ppARDTest, ppTSTest	phillips	phillips, unitroot	-	unitrootxcorr, x = 1-3
Stacionarita	KPSS	-	-	-	-	-	-
Kointegrace	CRDF	-	-	cadf, rztcrit	-	-	-
	CRADF	-	-	-	-	-	-
	CRDW	-	-	-	-	-	-
	Johansen	-	-	johansen, c_sja, c_sjt	-	-	coint
	POH	-	-	-	-	-	cointegrx, x = 1-3

Pozn.:

- ❑ POH = Phillips – Ouliaris – Hansenův test kointegrace,
- ❑ KPSS – test stacionarity autorů Kwiatkowski – Phillips – Shmidt – Shin [ARLT & ARLTOVÁ 2003], [BIERENS 2005], [COTTRELL 2005],
- ❑ **Stat** – Statistics Toolbox Matlabu (nej/novější verze – dle informací na webových stránkách firmy MathWorks),
- ❑ **GARCH** – toolbox Matlabu (nej/novější verze – dle informací na webových stránkách firmy MathWorks) pro modelování a analýzu volatility jednorozměrných časových řad. Výstupy testů jednotkových kořenů obsahují i p-value,
- ❑ **TS_STAT** – balík funkcí Time Series Statistics [KANZLER 1998], dostupný z <http://www2.gol.com/users/kanzler/index.htm#L.%20Kanzler:%20Software>,
- ❑ **ET** – Econometric Toolbox [LESAGE 1999],
- ❑ **TSA** – Time Series Analysis Toolbox [SCHLOEGL 1996],
- ❑ **NJ** – m-soubory přiložené k disertační práci Mgr. Jiřího Neubauera, Ph.D. [NEUBAUER 2005],
- ❑ DW statistiku obsahuje také balík LREGRESS (DW.m), autor: M.J.Chlond - Nov94, m.chlond@uclan.ac.uk.

Z přehledu je zřejmé, že žádný z dostupných balíčků neobsahuje všechny potřebné a vhodné funkce pro ekonometrii. Tyto funkce je možné pouze „poskládat“ z různých (i vlastních) zdrojů. Nejvíce funkcí obsahuje freeware Econometrics Toolbox, avšak jeho výsledky testování kointegrace jsou – vzhledem k dobře definovanému modelovanému a simulovanému systému – v podstatě nesprávné a nespolehlivé.

7 Závěr

Závěrem je možné konstatovat následující skutečnosti:

1. *Zdánlivá, klamná* korelace a regrese se vyskytuje pouze u *nestacionárních* veličin. Je způsobena nedodržením (neposouzením) základní podmínky stacionarity (ve střední hodnotě a rozptylu) vysvětlujících veličin při užití metody nejmenších čtverců.

2. *Povrchní analýza* pomocí „okometrického“, ale i statistického posouzení (klasické i moderní lineární regrese) *nekointegrovaných* veličin (1. řádu) vede ke klamným a nesmyslným závěrům s možným nepříznivým dopadem na interpretaci a realizaci výsledků v technické a technologické praxi.
3. Zdánlivou regresi může signalizovat *analýza reziduí*, a to hlavně výskyt jejich *pozitivní autokorelace* i *trendu*, ale hlavně splnění empirické *Grangerovy-Newboldovy* nerovnosti.
4. *Lineární regrese* poskytuje správné výsledky pouze u tzv. *kointegrovaných* nestacionárních veličin (prvního a vyššího řádu).
5. *Analýza kointegrace* pozůstává ze zjištění *integrovánosti* veličin a ze samotného testování výskytu *kointegrace*. Bývá však obsažena pouze v *ekonometrických* programech, či *toolboxech* – a ne v programech statistických, či v programech pro analýzu časových řad.
6. Jako vhodné (se správnými výsledky) *testy kointegrace* veličin simulovaného systému se jeví:
 - CRDW, CRDF a Johansenův test (tento však pouze v případě uvažování VAR(1) modelu s absolutním členem),
 - grafické posouzení charakteru reziduí,
 - kointegrační regresní rovnice ve tvarech modelů ADL a ECM.
7. Nespolehlivost (částečná nesprávnost) výsledků m-souborů testů jednotkových kořenů, kointegrace a jejich kritických hodnot ve freeware balíku *Econometrics Toolbox*, (jako i pro první část kointegrační analýzy v balíku *TS_STAT*) indikuje nutnost jejich korekcí, či vytvoření si m-funkcí vlastních. Pro samotnou firmu *MathWorks* to může být výzvou k vytvoření chybějícího, a přitom čím dále potřebnějšího, profesionálního toolboxu tohoto zaměření. Tento balík by navíc doplňoval kolekci obdobně zaměřených seskupení, mezi něž patří *Statistics*, *Financial* a *GARCH* toolbox.

Poděkování

Autor příspěvku tímto děkuje Mgr. Jiřímu Neubauerovi, Ph.D. z Univerzity obrany v Brně za poskytnutí jeho disertační práce v elektronické podobě, jako i za ověření výsledků testování kointegrace na prezentovaných simulovaných datech pomocí ním vytvořených m-funkcí v programu MATLAB.

Literatura

- [1] ARLT, J. 1997. *Regresní analýza nestacionárních ekonomických časových řad*. Politická ekonomie 45 (2), VŠE Praha, 1997, s. 281-289. ISSN 0032-3233.
- [2] ARLT, J. 1999. *Moderní metody modelování ekonomických časových řad*. 1.vyd. Praha: Grada Publishing, s.r.o., 1999. 312 s. ISBN 80-7169-539-4.
- [3] ARLT, J. & ARLTOVÁ, M. 2003. *Finanční časové řady*. 1. vyd. Praha : Grada Publishing, a.s., 2003, 220 s. ISBN 80-247-0330-0.
- [4] BENKOVSKIS, K. 2004. *Econometric models of Latvian imports*. Lithuania - Nordic Research Networking in Social Sciences 2003-2004. Dostupný z www: < URL: http://www.eurofakultetas.vu.lt/Nordic/Benkovski_Econometric%20models%20of%20Latvian%20imports.pdf >
- [5] BIERENS, H. 2005. *EeasyReg - Free Econometrics Software for Easy Regression Analysis* [online]. Pennsylvania State University, PA, 2005. Dostupný z www: < URL: <http://econ.la.psu.edu/~hbierens/> >
- [6] CIPRA, T. 1986. *Analýza časových řad s aplikacemi v ekonomii*. Praha : SNTL, 1986, 248 s.
- [7] COTTRELL, A. 2005. *Gnu Regression, Econometrics and Time-series Library - Gretl*. SW pro analýzu časových řad. Dostupný z www: < URL: <http://gretl.sourceforge.net/> >
- [8] CUMMINS, C. 2005. *Critical Values for the Durbin-Watson Test*. Dostupné z www: < URL: <http://www.stanford.edu/~clint/bench/dwcrit.htm> >
- [9] GAJDA, J. 2004. *Ekonometria*. 1. vyd. Warszawa : C.H.Beck, 2004. 306 s.

- [10] KANZLER, L. 1998. *Time Series Statistics TS_STAT*. Oxford : University of Oxford, Department of Economics, 1999. Dostupné z www: < URL: <http://users.ox.ac.uk/~econlrk> >
- [11] KUFEL, T. 2004. *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*. 1.vyd. Warszawa : Wydawnictwo Naukowe PWN, 2004. 154 s. ISBN 83-01-14284-7.
- [12] LESAGE, J. P. 1999. *Applied Econometrics using MATLAB*. Toledo : University of Toledo, Department of Economics, October 1999, 348 s. Manuál a Econometrics Toolbox jsou dostupné z www: < URL: <http://www.spatial-econometrics.com/> >
- [13] LUCCHETTI, R. 2005. *Appunti di analisi delle serie storiche*. Studijní materiál Dipartimento di Economia, Università Politecnica delle Marche, Ancona, Italy, 152 s. Dostupný z www: < URL: http://www.econ.unian.it/lucchetti/ects_en.html >
- [14] MACKINNON, J. G. 1996. *Numerical distribution functions for unit root and cointegration tests*. Journal of Applied Econometrics, 11, 1996, pp. 601-618. Program *urcdist.exe* dostupný z www: < URL: <http://qed.econ.queensu.ca/pub/faculty/mackinnon/> >
- [15] MARČEK, D. & MARČEK, M. 2001. *Analýza, modelovanie a prognózovanie časových radov s aplikáciami v ekonomike*. 1. vyd. Žilina : EDIS – vydavateľství Žilinské univerzity, 2001. 282 s. ISBN 80-7100-870-2.
- [16] MELOUN, M. & MILITKÝ, J. 1994. *Statistické zpracování experimentálních dat*. 1.vyd. Praha : PLUS, 1994. 839 s. ISBN 80-85297-56-6.
- [17] MELOUN, M. & MILITKÝ, J. 2004. *Statistická analýza experimentálních dat*. 2. upravené a rozšířené vyd. Praha : Academia, 2004. 953 s. ISBN 80-200-1254-0.
- [18] NEUBAUER, J. 2005. *Vybrané metody statistické analýzy náhodných procesů a jejich aplikace*. Disertační práce. Ostrava : Ostravská univerzita, Přírodovědecká fakulta, katedra matematiky, 2005. 123 s.
- [19] SHUMWAY, R. H. & STOFFER, D. S. 2000. *ASTSA – Applied Statistical Time Series Analysis*. SW pro aplikovanou statistickou analýzu časových řad. Dostupný z www: < URL: <http://www.stat.ucdavis.edu/~shumway/tsa.html> >
- [20] SCHLOEGL, A. 1996. *Time Series Analysis Toolbox TSA*. Graz : University of Technology, Department of Medical Informatics, 1996. Dostupné z www: < URL: <http://www.dpmi.tu-graz.ac.at/~schloegl/lectures/TSA/index.htm> >
- [21] VESELÝ, V. 2003. *Úvod do časových řad*. In sborník celostátního semináře ANALÝZA DAT 2003/II. Pardubice : Trilobyte, listopad 2003, s. 7–31. ISBN 80-239-2590-3.

Ing. Jan Morávka, Ph.D.

739 61 Třinec – Staré město, Frýdecká 126, e-mail: jan.moravka@tzi.trz.cz, tel.: 558 53 2192