Základní transformace časových řad

Veškeré nástroje základní korelační analýzy, kam patří i lineární regresní (ekonometrické) modely a modely časových řad jsou ve své podstatě založeny na dvou předpokladech

- linearita zkoumaného vztahu,
- existence konstantní střední hodnoty zkoumaných (vysvětlovaných i vysvětlujících) časových řad – resp. slabou stacionaritu těchto časových řad, která je definována ještě dalšími podmínkami.

Charakter většiny vztahů mezi ekonomickými veličinami a charakter časových řad těchto ekonomických veličin však tyto předpoklady nesplňují, proto je k tomu, aby byla analýza smysluplná, nutné vybrat a provést některé elementární transformace. Většinou existuje několik možností, jejich volba závisí spíše na ekonomické teorii než statistické teorii samotné.

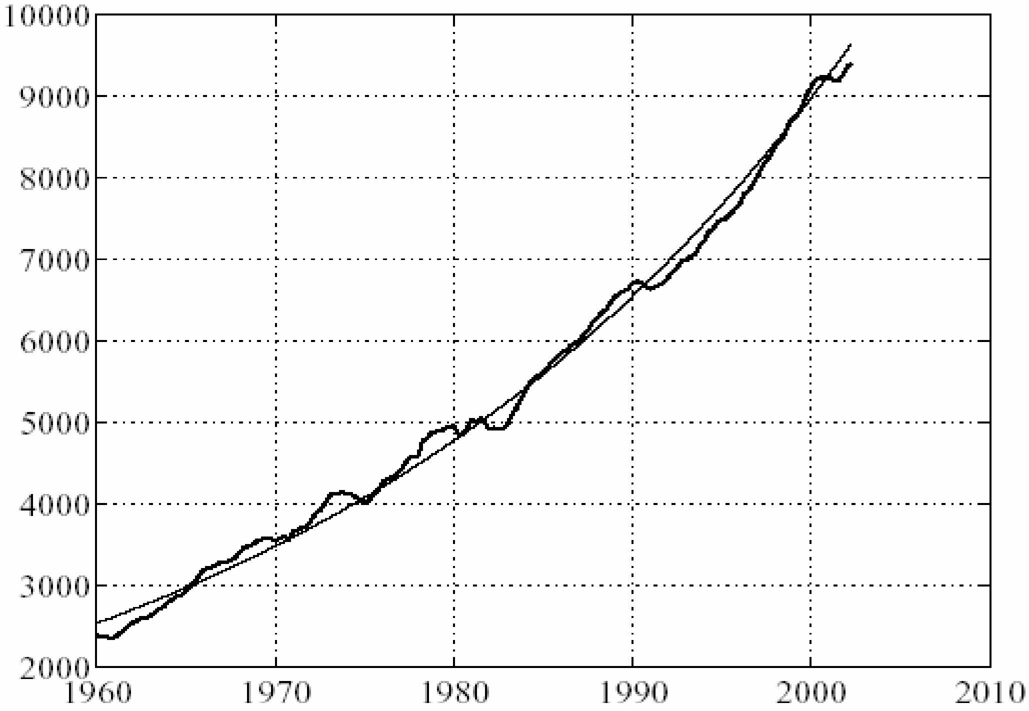
Linearita zkoumaných vztahů. Většinu závislostí a procesů v ekonomii lze popsat spíše pomocí procentních změn než pomocí změn v absolutních hodnotách daných veličin. Uvedeme dva typické příklady:

1. Vztah dvou veličin – např. jednoduchá rovnice poptávky po penězích, kdy odhlédneme od vlivu alternativních nákladů (nominální úrokové míry) a budeme předpokládat, že poptávané množství reálných peněz je determinováno reálným množstvím transakcí, které můžeme měřit např. reálnou spotřebou (tj. reálnou hodnotou nákupů) domácností. Jestliže budeme dále předpokládat, že jednoprocentní změna reálné spotřeby vyvolá α -procentní změnu poptávky po reálných penězích, má takový model jednoduchý tvar
$$\frac{M_t}{P_t} = \mathrm{M}_t = kC_t^{\alpha} \qquad (1)$$
kde k je úrovňová konstanta, která kvalitu vztahu mezi oběma veličinami (reálnými penězi $\mathrm{M}_t = M/P$ a reálnou spotřebou C) absolutně neovlivní. Skutečnost, že jednoprocentní změna C vyvolá α -procentní změnu poptávky po reálných penězích, lze ukázat jednoduše pomocí diferenciálního počtu
$$\frac{\frac{\partial \mathrm{M}_t}{\mathrm{M}_t}}{\frac{\partial C_t}{C_t}} = \frac{\partial \mathrm{M}_t}{\partial C_t}\frac{C_t}{\mathrm{M}_t} = \alpha k C_t^{\alpha-1}\frac{C_t}{\mathrm{M}_t} = \alpha k C_t^{\alpha-1}\frac{C_t}{kC_t^{\alpha}} = \alpha$$

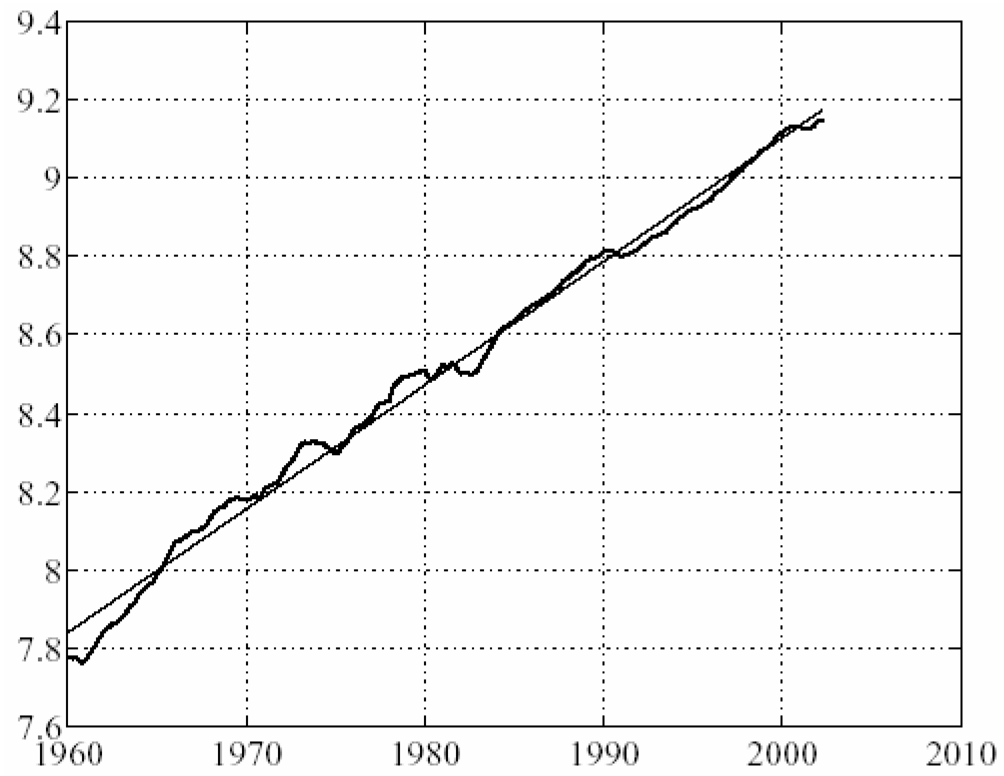
2. Vývoj jedné veličiny v čase – u ekonomických časových řad má většinou smysl mluvit o procentním tempu růstu (třeba i časově proměnném) ekonomických veličin, nikoliv o růstech vyjádřených absolutním přírůstkem
$$Y_t = Y_{t-1}\left(1+g_t\right).$$

Oba dva uvedené příklady vedly k nelineárnímu popisu chování. Linearizace těchto příkladů je však velmi triviální (a je naštěstí aplikovatelná ve velké části moderní makroekonomie) – stačí je zlogaritmovat a dostáváme
$$\log \mathrm{M}_t = \log k + \alpha \log C_t \qquad \text{resp.} \qquad \log Y_t = \log Y_{t-1} + g_t .$$
Vizuálně si můžeme dopady logaritmu na charakter časové řady ukázat na příkladu reálného HDP (USA).

Obr. 1 – USA: reálný HDP – exponenciální trend



Obr. 2 - USA : logaritmus reálného HDP - lineární trend

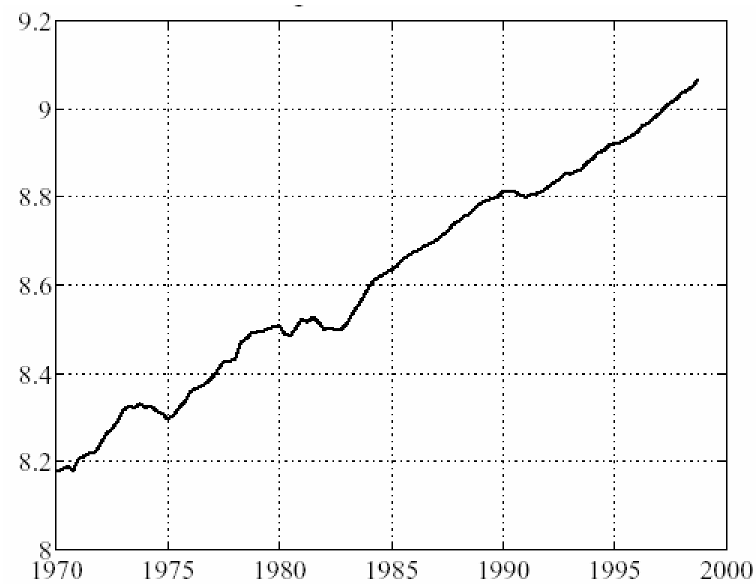


Na obr.1 je původní časová řada reálného HDP s odhadnutým jednoduchým exponenciálním trendem, na obr. 2 je zlogaritmovaná časová řada HDP, která má zcela zjevně charakter lineární, nikoliv exponenciálního růstu (naznačen je tudíž odhad lineárního trendu).

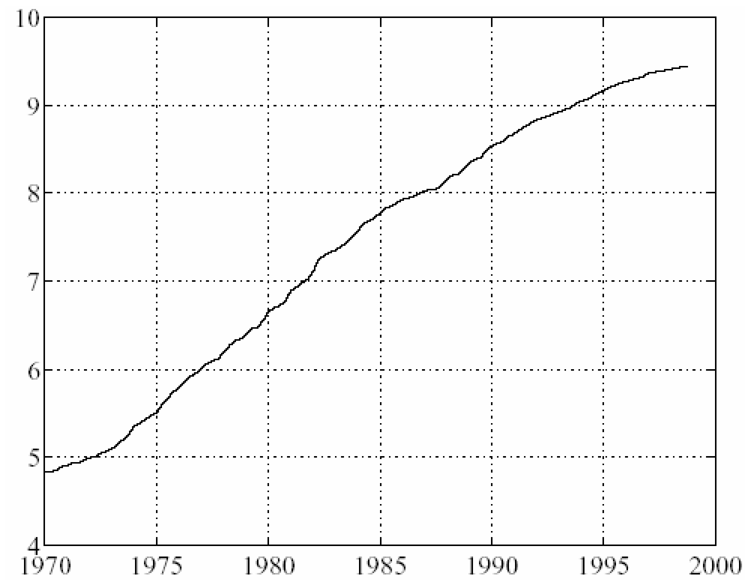
Stacionarita časové řady. Provádět jakoukoli analýzu založenou na korelaci (kam patří i regresní modely či modely časových řad) musíme velmi opatrně, neboť se jinak zcela jednoduše můžeme dostat k absolutně nesmyslným výsledkům.

Problém můžeme ilustrovat velmi jednoduchým příkladem. Vezmeme dvě ekonomické nestacionární řady, které spolu zcela zjevně absolutně nesouvisí, reálné HDP USA (obr.3) a index nominálních mezd v Řecku (obr.4), obě dvě na období 1970-1998.

Obr. 3 - USA : log of real GDP



Obr. 4 - Greece : Log of nominal wage rate



Vypočítáme-li koeficient korelace těchto dvou řad, dostaneme 0.96. Podobně, jestliže odhadneme jednoduchý regresní model ve tvaru

$$\log Y_t = \alpha + \beta \log W_t + \varepsilon_t$$

Dostaneme odhad β 5.92se směrodatnou chybou odhadu 0.09 a tedy t-statistikou (t-poměrem) 64.38.

Oba dva výsledky by měly ukazovat na velmi silnou vazbu mezi oběma veličinami. Oba výsledky jsou však zjevně naprosto nesmyslné – příčina této nesmyslnosti spočívá právě v nestacionaritě obou řad. Koeficient korelace se totiž velmi zjednodušeně počíta tak, že vezmeme u obou časových řad jejich průměr (zde bude samozřejmě ležet někde uprostřed časové řady), vypočítáme odchylky dané časové řady od tohoto průměru v každém období, a srovnáváme, jestli velkým kladným odchylkám v jedné řadě odpovídají ve stejném období velké kladné odchylky v druhé řadě (pozitivní korelace), nebo popř. velké záporné odchylky v druhé řadě (negativní korelace). Čím více platí toto pravidlo, tím silnější korelace, a tím blíže je koeficient korelace k 1 (pozitivní korelace), popř. k -1(negativní korelace).

Jestliže však tedy máme dvě časové řady, které systematicky rostou, budou mít obě vždy na svém začátku velké záporné odchylky a na svém konci velké kladné odchylky od svých průměrů (už jenom počítat průměr u nestacionárních, tj. systematicky rostoucích časových řad je samozřejmě nesmysl), a tedy musíme zákonitě dostat vysokou míru korelace, resp. velmi významný odhad parametrů v regresním modelu.

Existuje několik řešení, výběr z těchto možností je, jak už bylo řečeno, spíše otázkou ekonomické teorie než statistiky samotné.

1) Stacionarizace časové řady
- jejím diferencováním,
- odhadnutím jejího trendu a prací s odchylkou této řady od trendu, tj. tzv. filtrací.

Tyto typy analýz (tj. analýzy na stacionarizovaných časových řadách) se samozřejmě hodí spíše pro zkoumání cyklických, tj. ne příliš dlouhodobých vlastností časových řad. Někdy se však zajímáme právě o dlouhodobé vztahy mezi ekonomickými veličinami, které můžeme modelovat pouze poněkud komplikovanějším způsobem – v tzv. error-correction tvaru.

2) Modelování ve tvaru error-correction (bohužel neexistuje žádný zažitý uznávaný český překlad) umožňuje konkrétní odhad dlouhodobých vztahů mezi diferencovanými nebo nefiltrovanými časovými řadami (korektní v tom smyslu, že se na něj nevztahuje výše uvedená výtka ohledně nestacionarity).

Diferencování časových řad. Diferencováním logaritmu časové řady

$$\Delta \log P_t = \log P_t - \log P_{t-1}$$

ve skutečnosti získáme aproximaci tempa jejího růstu a tento typ stacionarizace má tedy velmi jednoduchou ekonomickou interpretaci.

Jediným problémem je volba typu diferenciace ve smyslu délky zpoždění, se kterým diferenci počítáme – tj. zda počítáme rozdíl současné hodnoty a např. hodnoty v minulém čtvrtletí (tzn. mezikvartální změny) nebo hodnoty ve stejném čtvrtletí minulého roku (tzn. meziroční změny), atd.

V naprosté většině případů však neexistuje žádný důvod používat k analytickým účelům jiné než mezikvartální, resp. meziměsíční diference (časovou řadu diferencujeme vůči nejbližšímu předchozímu období). Meziroční diference se používají pouze pro rychlé získání povrchní vizuální představy o tom, co se v ekonomice zhruba dělo, neboť bývají méně volatilní a tudíž o něco více „čitelné“ než mezikvartální nebo meziměsíční změny.

Časová řada meziročních změn totiž v sobě obsahuje uměle vytvořenou autokorelaci, která může významně zkreslit výsledky další kvantitativní analýzy. Opět použijeme jednoduchý příklad.

Nasimulujeme tzv. náhodnou procházku

$$x_t = x_{t-1} + \varepsilon_t$$

kde ε_t je bílý šum (důležitý je zde předpoklad nulové autokorelace).

První diferenci této řady je samozřejmě přímo ε_t

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_t$$

a tudíž z definice neobsahuje žádnou autokorelaci,

$$\operatorname{cov}(\Delta x_t, \Delta x_{t-1}) = \operatorname{cov}(\varepsilon_t, \varepsilon_{t-1}) = 0$$

Naopak, vytvoříme-li časovou řadu diferencií přes 4 období (jakoby „meziroční“ změny), pokud bude x_t p.a.

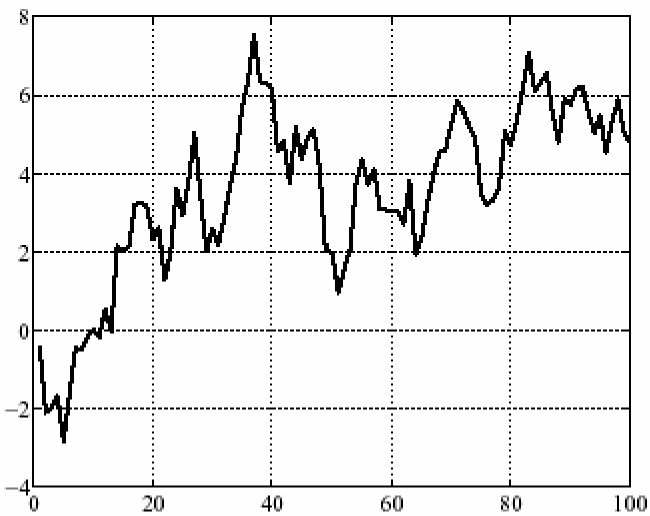
$$\Delta_4 x_t = x_t - x_{t-4} = \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3}$$

je zřejmé, že bude autokorelována, neboť

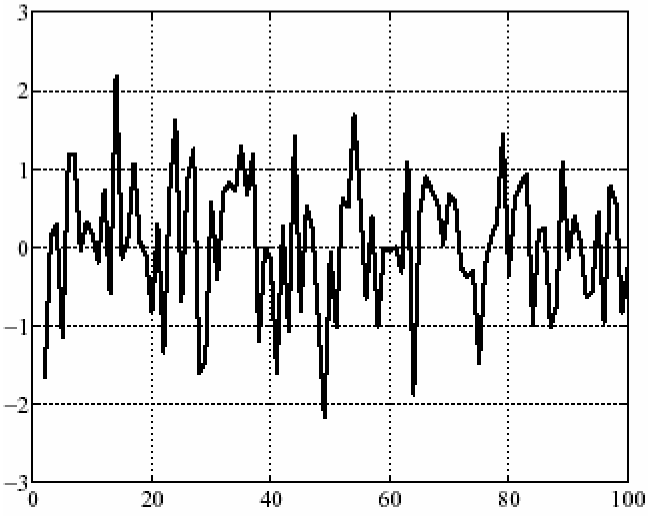
$$\operatorname{cov}(\Delta_4 x_t, \Delta_4 x_{t-1}) = \operatorname{cov}(\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3}, \varepsilon_{t-1} + \varepsilon_{t-2} + \varepsilon_{t-3} + \varepsilon_{t-4}) = 3\sigma_\varepsilon^2 \qquad (2)$$

Kde σ_ε^2 je rozptyl náhodné složky ε. Na obr.5, 6, 7 uvádíme původní nasimulovanou časovou řadu x_t, a její diference $\Delta x_t = x_t - x_{t-1}$ a $\Delta_4 x_t = x_t - x_{t-4}$. Autokorelovanost poslední časové řady je vizuálně zřejmá i bez kvantitativních testů.

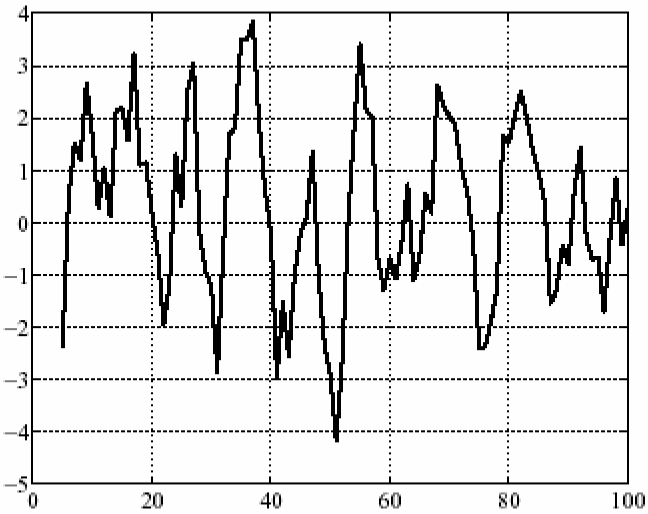
Obr. 5 - Simulovaná časová řada $x_t = x_{t-1} + \varepsilon_t$



Obr. 6 - 1.diference simulované časové řady $\Delta x_t = x_t - x_{t-1} = \varepsilon_t$



Obr. 7 - Meziroční diference simulované časové řady $\Delta_4 x_t = x_t - x_{t-4}$



Filtrace časových řad. Filtrací zde budeme rozumět specificky pouze rozklad dané časové řady na trendovou složku a na cyklickou složku. Nástroje na tento typ rozkladu jsou v principu jednorozměrné, tj. využívají pouze jednu danou časovou řadu, nebo vícerozměrné (strukturální), které k extrakci trendu z dané časové řady využívají informaci i z jiných časových řad (např. pro odhad trendu a mezery HDP můžeme využít informaci z pohybu inflace, reálných úrokových měr, apod.).

Jako příklad uvedeme jeden z nejčastěji používaných jednorozměrných filtrů, tzv. Hodrick-Prescottův (HP) filtr. Je nutné si uvědomit, že HP filtr je smysluplné používat pouze na logaritmovaných časových řadách, tak aby první diference měla význam tempa růstu.

$$x_t = \bar{x}_t + \hat{x}_t$$

$$\Delta\bar{x}_t = \Delta\bar{x}_{t-1} + \varepsilon_t$$

$$\hat{x}_t = \omega_t$$

kde x_t je původní časová řada, $\bar{x}_t$ je její trend, $\hat{x}_t$ je její mezera (cyklická složka), ε_t je šok (náhodná porucha) v tempu růstu trendu, ω_t je cyklický šok (oba dva předpokládané bílé šumy).

HP má tedy dvě význačné charakteristiky:

1. Dostatečná „hladkost“ trendu je zajištěna tím, že šoky nepůsobí přímo na úroveň trendu (na $\bar{x}_t$), nýbrž na jeho tempo růstu $\Delta\bar{x}_t$.
2. Cyklická část je modelována jako pouhý bílý šum, což se může zdát nerealistické, v praxi však toto omezení není příliš limitující a bude déle objasněno na cvičeních.

Při použití HP filtru musíme jako jediný volný parametr zadat poměr rozptylu cyklického šoku ku rozptylu trendového šoku,

$$\frac{\sigma_\omega^2}{\sigma_\varepsilon^2}.$$

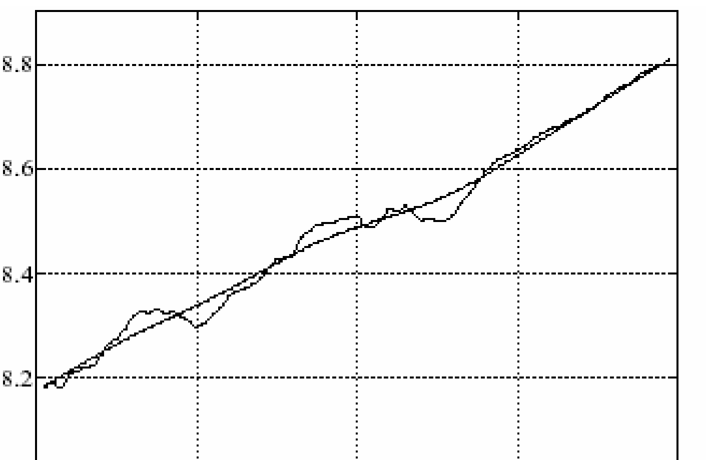
Platí, že čím nižší je tento parametr, tím nižší část variability dané původní časové řady je považována za cyklické kolísání a vyšší část je považována za změny trendu (a naopak). Jinými slovy, čím nižší je tento parametr, tím těsněji výsledný trend sleduje původní řadu.

Autoři filtru provedli výzkum širokého spektra časových řad a zkoumali vlastnosti trendových a cyklických složek různých ekonomických časových řad s využitím standardních moderních teorií ekonomického růstu a ekonomického cyklu a jejich hrubé doporučení je používat

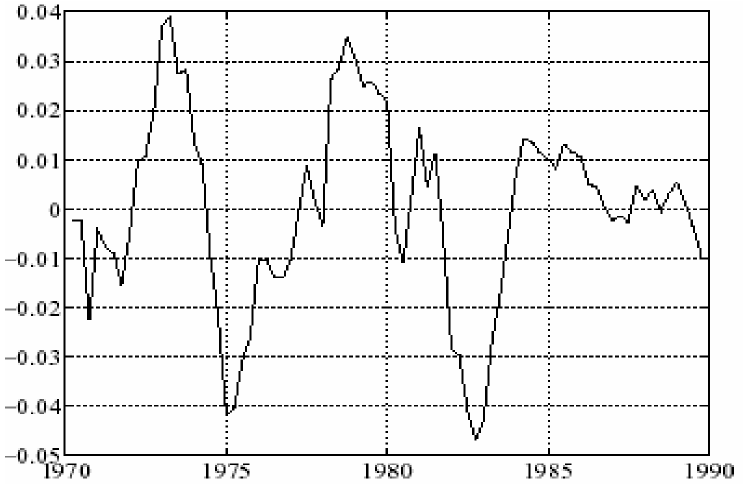
- pro roční časové řady 100
- pro čtvrtletní časové řady 1600
- pro měsíční časové řady 14400.

Na obr. 8 a 9 je pro ilustraci trend a mezera (logaritmu) reálného HDP USA (čtvrtletní časová řada) s parametrem 1600 (tj. v souladu s doporučením Hodricka a Prescotta), na obr.10 a 11 je totéž s parametrem 100.

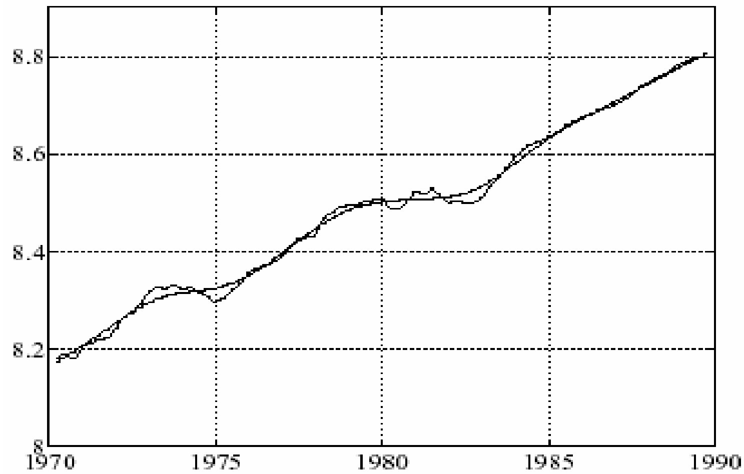
Obr. 8 - USA : krátkodobý kvartální reální HDP, HP filtr: $\frac{\sigma_\omega^2}{\sigma_\varepsilon^2}$ **=1600**



Obr. 9 - USA : cyklická část reálného HDP, HP filtr: $\frac{\sigma_{\omega}^{2}}{\sigma_{\varepsilon}^{2}}$ **=1600**



Obr. 10 - USA : krátkodobý kvartální reální HDP, HP filtr: $\frac{\sigma_{\omega}^{2}}{\sigma_{\varepsilon}^{2}}$ **=100**



Obr. 11 - USA : cyklická část reálneho HDP, HP filtr: $\frac{\sigma_{\omega}^{2}}{\sigma_{\varepsilon}^{2}}$ **=100**