Testování dlouhodobých vztahů (error-correction)

Pojem „dlouhodobý“ vztah mezi veličinami si definujeme následujícím velmi zjednodušeným způsobem: Mějme několik veličin, pro lepší srozumitelnost řekneme konkrétně logaritmus množství peněz (tj. nějaký peněžní agregát) m_t, logaritmus cenové hladiny p_t, a logaritmus reálného výstupu y_t. Všechny tři časové řady mohou být obecně nestacionární. Jestliže se zajímáme o to, zda v dlouhém období (tj. ne v každém okamžiku, ale vždy „v průměru“ za dostatečně dlouhou dobu) platí např.

$$m_t = \alpha + \beta p_t + \gamma y_t \qquad (1)$$

Tak jinými slovy hledáme takové α, β, γ aby

$$m_t - \alpha - \beta p_t - \gamma y_t \qquad (2)$$

byla stacionární časová řada s nulovou střední hodnotou. Tím říkáme, že se sice mohou vyskytnout odchylky od výše uvedeného předpokládaného vztahu (1), ale tyto odchylky fluktuují kolem nuly, jsou krátkodobé a v dlouhém období „vymizí“.

Odhad dlouhodobého vztahu (1) není jednoduchý, neboť

1. pokud jsou m_t, p_t, y_t nestacionární, nemůžeme použít obyčejnou regresi
$$m_t = \alpha + \beta p_t + \gamma y_t + \varepsilon_t$$
z důvodů zmíněných v minulé přednášce;
2. pokud bychom odhadovali analogickou rovnici s použitím prvních diferencí
$$\Delta m_t = \alpha + \beta \Delta p_t + \gamma \Delta y_t + \varepsilon_t$$
nemohli bychom samozřejmě dělat vůbec žádné závěry o existenci dlouhodobých vztahů mezi původními úrovněmi (nikoliv diferencemi) těchto veličin. To, že existuje neměnný těsný vztah např. mezi inflací a růstem množství peněz nevypovídá vůbec nic o tom, že existuje neměnný dlouhodobý vztah mezi cenovou hladinou a množstvím peněz.

Odhad testování tohoto typu dlouhodobých vztahů je poměrně komplikovaná záležitost a spadá do tzv. kointegrační analýzy. Nicméně při jednom omezujícím předpokladu lze celou kointegrační analýzu převést na jednoduchý regresní model. Předpokládáme-li, že vždy když se vyskytne z jakéhokoliv důvodu odchylka od (1), bude vždy korigována pohybem pouze jedné a vždy té stejné veličiny (např. pohybem množství peněz m_t) a v ostatních veličinách se existence této odchylky od dlouhodobého vztahu neprojeví (tehdy říkáme, že tyto veličiny, tj. p_t, y_t jsou vzhledem k m_t tzv. slabě exogenní), můžeme dlouhodobý vztah (1) odhadovat v následujícím tvaru

$$\Delta m_t = \kappa\left(m_{t-1} - \alpha - \beta p_{t-1} - \gamma y_{t-1}\right) + \cdots + \varepsilon_t \qquad (3)$$

kde κ je záporná konstanta (pokud bychom získali odhad této konstanty kladný, nemůžeme mluvit o existenci dlouhodobého vztahu, viz. níže interpretace této rovnice), a místo tří teček … se dosadí další vysvětlující proměnné, které neovlivní dlouhodobý vztah, ale přispějí k vysvětlení krátkodobých fluktuací (odchylek od tohoto dlouhodobého vztahu). Nejčastěji užívanými kandidáty na tyto další vysvětlující proměnné jsou zpožděné diference Δm_{t-1}, Δm_{t-2}, adt., popř. Δp_{t-1}, Δy_{t-1} atd. Jejich výběr je dán jednak ekonomickou teorií (která ovšem ve většině případů neposkytuje v tomto typu modelu jasné vodítko) a jednak statistické vlastnosti modelu, tj. např. vlastnosti, které požadujeme na reziduích apod. K čistě statistickému výběru těchto vysvětlujících proměnných lze použít tzv. informační kritéria, viz. níže.

Intuitivní interpretace rovnice (3) je poměrně jednoduchá: Vyskytnuli se v období *t-1*odchylka od dlouhodobého vztahu, znamená to jinými slovy, že

$$m_{t-1} - \alpha - \beta p_{t-1} - \gamma y_{t-1} \neq 0$$

V případe >0 můžeme tuto situaci popsat jako „příliš vysoké“ m_{t-1} (příliš mnoho peněz v oběhu) vzhledem k daným p_{t-1}, y_{t-1} (tento výrok je samozřejmě podmíněn výše uvedeným předpokladem exogenity p_t a y_t - jinak bychom nemohli rozlišit, jestli náhodou není naopak p_{t-1} příliš nízké vzhledem k m_{t-1} a y_{t-1}, nebo y_{t-1} příliš nízké vzhledem k m_{t-1} a p_{t-1}). Vzniklá kladná odchylka se tak v rovnici (3) poté násobí zápornou konstantou κ a celý člen

$$\kappa\left(m_{t-1} - \alpha - \beta p_{t-1} - \gamma y_{t-1}\right)$$

(záporný) způsobuje pokles Δm_t v následujícím období *t*. Množství peněz je tudíž korigováno směrem k obnovení dlouhodobého vztahu (proto se tvar (3) nazývá error-correction).
V případě <0 platí přirozeně analogický sled opačných událostí.

Informační kritéria. Informační kritéria se obecně používají při výběru „nejlepšího“ tvaru regresního modelu tam, kde nelze „nejlepší“ tvar určit čistě z ekonomické teorie. V našem konkrétnim případě bychom měli vybrat tu „nejlepší“ z různých možností, jaké vysvětlující proměnné přidat místo tří teček… v rovnici (3).

Problém je v tom, že „nejlepší“ model musí ze statistického hlediska splňovat dva vzájemně protikladné požadavky :

1. co nejlepší vysvětlovací schopnost, tj. schopnost vysvětlit co největší část pohybu vysvětlované proměnné pomocí vysvětlujících proměnných a co nejmenší část pohybů přisoudit náhodným vlivům (reziduím),
2. jednoduchost modelu ve smyslu použití co nemenšího počtu vysvětlujících proměnných (nalezení těch „správných“ vysvětlujících proměnných). Problém totiž je, že můžeme postavit model, který bude mít 100% vysvětlovací schopnost – stačí do něj zahrnout tolik vysvětlujících proměnných, jaký je počet pozorování – takový model však nebude mít žádnou vypovídací schopnost a nebude robustní : abychom udrželi danou vysvětlovací schopnost modelu, museli bychom ho měnit s každým dalším pozorováním, které bychom do modelu dodali.

Informační kritéria obecně jistým způsobem porovnávají/váží vysvětlovací schopnost modelu a počet vysvětlujících proměnných do modelu zahrnutých. Informačních kritérií existuje několik (které z nich použijeme záleží spíše jen na nás), používají se však v principu stejným způsobem – nejdříve odhadneme všechny modely, z nichž chceme vybrat „nejlepší“ verzi, pro každý z těchto modelů vypočítáme níže uvedené kritérium a jako „nejlepší“ vybereme ten model, u kterého je hodnota kritéria nejnižší ze všech.

1. Akaikeho informační kritérium (AIC)
$$AIC = \log \sigma_\varepsilon^2 + \frac{2}{T} k$$
2. Schwartzovo bayesiánské informační kritérium (SBC)
$$SBC = \log \sigma_\varepsilon^2 + \frac{\log T}{T} k$$

kde v obou případech σ_{ε}^{2} je odhad směrodatné chyby reziduí modelu, T je počet pozorování, na kterých byl model odhadnut, k je počet vysvětlujících proměnných (včetně konstanty).

Na první pohled je zřejmé, že kritérium bude tím nižší (tj. tím „lepší“), čím je nižší směrodatná chyba reziduí (tedy čím je větší vysvětlovací schopnost modelu) a čím je nižší počet vysvětlujících proměnných zahrnutých do modelu.

Jednou z možných strategií pro výběr „krátkodobých“ vysvětlujících proměnných v modelu (3) namísto tří teček … je např. odhadnout postupně několik modelů, do kterých postupně zahrneme

$$
\begin{aligned}
&\Delta m_{t-1} \\
&\Delta m_{t-1} + \Delta m_{t-2} \\
&\vdots \\
&\Delta m_{t-1} + \Delta m_{t-2} + \cdots + \Delta m_{t-n}
\end{aligned}
$$

kde n je „rozumně“ zvolené číslo (je např. nesmyslné předpokládat, že by na současný pohyb množství peněz měly vliv všechna jeho zpoždění až do doby dva nebo více let zpět – se čtvrtletními časovými řadami můžeme tudíž položit $n = 8$ nebo dokonce i méně), a z takto odhadnuté sady modelů vybereme ten, který minimalizuje AIC nebo SBC.

Jinou variantou je samozřejmě přidávat postupně nejen Δm_{t-j}, ale současně s tím také Δp_{t-j} a Δy_{t-j}. Pokud nás zajímá opravdu pouze odhad a testování dlouhodobého vztahu a nikoliv proces krátkodobé dynamiky odchylek od tohoto vztahu, neměl by být mezi těmito a podobnými variantami statisticky významný rozdíl.