Možnosti testování sezonních jednotkových kořenů demografických časových řad v systému GRETL

The possibilities in testing of seasonal unit roots in demographic time series with GRETL system

Ondřej Šimpach, Petra Dotlačilová, Jitka Langhamrová

Abstract: The aim of this study is to present the options for testing seasonal unit roots in quarterly published demographic time series, in which the presence of seasonality is generally expected. The testing using sophisticated HEGY test will be demonstrated in an econometric package GRETL and with selected demographic time series with quarterly frequency the absence of stationarity will be proved.

Abstrakt: Předkládaná studie představí možnosti testování sezonních jednotkových kořenů u čtvrtletně publikovaných demografických časových řad, u nichž je obecně přítomnost sezonnosti očekávána. Testování s využitím sofistikovaného HEGY testu bude demonstrováno v prostředí ekonometrického balíčku GRETL a na vybraných demografických časových řadách s čtvrtletní frekvencí bude nepřítomnost stacionarity prokázána.

Key words: seasonal unit roots, HEGY test, demographic time series

Klíčová slova: sezonní jednotkové kořeny, HEGY test, demografické časové řady

JEL classification: C22, C52

Úvod

Pro potřeby demografické analýzy je zapotřebí, aby demografická časová řada měla mimo jiné i předpoklad nestacionarity, neboli aby se v příslušné časové řadě vyskytoval trend. Právě přítomnost trendu je jednou z mnoha podmínek, že analyzovaná demografická časová řada může být využita pro modelování a případné predikce. Předkládaný článek pojedná o možnostech exaktního určení, zda příslušná demografická časová řada, publikovaná s čtvrtletní frekvencí, je nebo není stacionární. Obecně neplatí, že každá čtvrtletně publikovaná časová řada je sezonní (viz Arlt, Arltová [1]). Pravděpodobnost, že čtvrtletní demografická časová řada sezonní je, je ovšem vysoká, proto bude představen přístup hledání sezonních jednotkových kořenů v čtvrtletně publikovaných demografických časových řadách, poprvé představen Hyllebergem et al. [6]. Ověření stacionarity je jednou z podmínek, aby s časovou řadou mohlo být následně nakládáno např. dle přístupu autorů Boxe a Jenkinse [3] pro modelování časových řad. Ve zvolených demografických časových řadách figurují *počty sňatků, počty rozvodů, počty živě narozených osob, počty potratů celkem, počty zemřelých celkem, počty přistěhovalých a počty vystěhovalých* a byly publikovány Českým statistickým úřadem (ČSÚ).

1. Metodika

Při pohledu na průběh ekonomické či demografické časové řady je možné vyslovit hypotézu o stacionaritě časové řady. V případě, že trend roste, respektive kolísá a má cykly, je možno tvrdit, že konkrétní časová řada je nestacionární. Pak také autokorelační funkce (ACF) má svou první hodnotu velmi vysokou, blížící se k jedné a ostatní zbylé hodnoty klesají jen velmi pomalu. Grafické zobrazení a definici ACF podává např. Arlt et. al v [2]. V případě, že je časová řada nesezónní a je třeba stacionaritu ověřit exaktně, je možno použít tzv. testů jednotkových kořenů, sestavených Dickeyem a Fullerem v [4]. K následujícím modelům:

a) $X_t = \Phi X_{t-1} + a_t$

b) $X_t = c + \Phi X_{t-1} + a_t$

c) $X_t = c + \Upsilon_t + \Phi X_{t-1} + a_t$

kde X_t je konkrétní časová řada, c je konstanta a a_t je reziduum, se vyslovuje hypotéza:

H_0: $\Phi = 1$, tj. časová řada je typu I(1),

H_1: $\Phi < 1$, je typu I(0).

Použité testové kritérium

$$T = \frac{\widehat{\Phi} - 1}{S_{\widehat{\Phi}}} \tag{1}$$

má t-rozdělení. Však vzhledem k tomu, že testovaná hypotéza je „nestacionarita“, nemá statistika t standardní t rozdělení, ale rozdělení, které bylo odvozené Dickeyem a Fullerem např. v [4]. Ve prospěch alternativní hypotézy svědčí nízké hodnoty testového kritéria t, respektive hodnoty nižší než α procentní kvantil rozdělení D-F. V případě, že v modelech a) b) a c) je zbytková složka autokorelovaná, potom se konstruuje tzv. rozšířený D-F test, který se liší pouze tím, že modely rozšíří o složku $\sum_{i=1}^{p-1} \Upsilon\, \Delta X_{t-i}$, takže

a) $\Delta X_t = \Phi X_{t-1} + \sum_{i=1}^{p-1} \Upsilon\, \Delta X_{t-i} + a_t$

b) $\Delta X_t = c + \Phi X_{t-1} + \sum_{i=1}^{p-1} \Upsilon\, \Delta X_{t-i} + a_t$

c) $\Delta X_t = c + \Upsilon_t + \Phi X_{t-1} + \sum_{i=1}^{p-1} \Upsilon\, \Delta X_{t-i} + a_t$

a $\Delta X_{t-i} = X_{t-1} - X_{t-i-1}$ je rozdíl sousedních hodnot.

V případě, že uvažujeme sezónní časovou řadu, nelze použít přístup testu jednotkového kořene autorů Dickey-Fuller. V současné době je možno použít přístupu, sestaveného autory Hylleberg et. al v [6], označeném jako HEGY, který byl speciálně sestaven pro testování přítomnosti sezónních jednotkových kořenů ve čtvrtletně pozorovaných časových řadách Y_t. Vše spočívá v testování statistické významnosti parametru π_i, kde $i = 1, \dots, 4$ v regresní rovnici, která dle Harveyho a Dijka [7] může mít následnou podobu:

$$\Delta_4 y_t = \mu_t + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-1} + \pi_4 y_{3,t-1} + \sum_{j=1}^{k} \Phi_j \Delta_4 y_{t-j} + \varepsilon_t \tag{2}$$

kde *t* = 1, … , *T*. Zároveň mějme Δ_k představující tzv. filtr, který může být definován jako

$$\Delta_k y_t \equiv (1 - B^k) y_t \equiv y_t - y_{t-k} \qquad \forall\, k = 1,2, \dots \tag{3}$$

kde B je operátor zpoždění. V regresní rovnici představuje μ_t deterministický trend. Nyní nechť

$$y_{1,t} = (1 + B + B^2 + B^3) y_t \tag{4}$$

$$y_{2,t} = -(1 + B + B^2 + B^3) y_t \tag{5}$$

$$y_{3,t} = -(1 - B^2) y_t \tag{6}$$

a vzhledem k tomu, že $(1 - B^4) = (1 - B) \cdot (1 + B) \cdot (1 + B^2)$, pak y_t může obsahovat sezónní jednotkové kořeny. Uvedené filtry, vedoucí k $y_{1,t}$, $y_{2,t}$ a $y_{3,t}$ odstraňují všechny jednotkové kořeny, kromě jednoho, který vyplývá ze skutečnosti, že roční filtr $(1 - B^4)$ může být rozložen jako

$(1 - B^4) = (1 + B + B^2 + B^3) \cdot (1 - B)$ nebo

$(1 - B^4) = -(1 + B + B^2 + B^3) \cdot (1 + B)$ nebo

$(1 - B^4) = -(1 - B^2) \cdot (1 + B^2)$.

Když se v regresní rovnici (2) parametr

- $\pi_1 = 0$, pak rovnice y_t obsahuje nesezónní jednotkový kořen,
- $\pi_2 = 0$, pak rovnice y_t obsahuje sezónní jednotkový kořen v pololetních četnostech, zmenšených o jednotku,
- $\pi_3 = \pi_4 = 0$, pak rovnice y_t obsahuje sezónní jednotkové kořeny v ročních četnostech $\pm i$, kde $i = 1, \dots, 4$.

Autoři Hylleberg et. al v [6] doporučují použití jednostranných klasických t-testů k určení statistické významnosti parametrů π_1 a π_2 a dále F-testu pro společnou statistickou významnost parametrů π_3 a π_4. V ekonometrickém systému Gretl je možno využít doplňku HEGY pro testování jednotkových kořenů v sezónních řadách. Testovaná hypotéza

H_0: časová řada je stacionární

H_1: non H_0

je zde vyjádřena jako

H_0: parametr z_1 ; z_2 ; z_3 ; $z_4 \neq 0$

H_1: non H_0

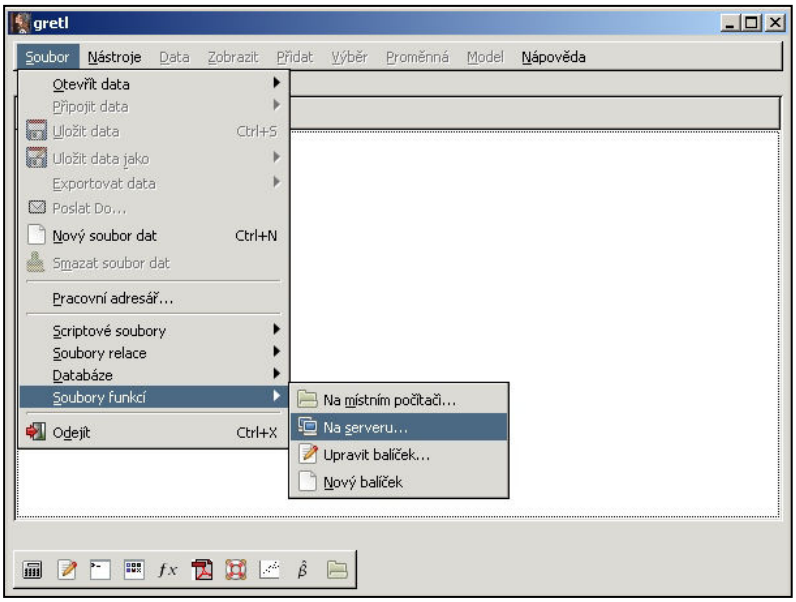
Oproti přístupu autorů Dickeyho a Fullera, kde nulová hypotéza znamená nestacionaritu, je v tomto případě testovaná hypotéza stacionarita. Systém Gretl má parametry π_1 ; π_2 ; π_3 a π_4 označeny jako z_1 ; z_2 ; z_3 a z_4 .

2. Instalace doplňku HEGY do systému GRETL

Bude výhodné ukázat postup, jak doplněk HEGY do systému získat, neboť ověřování stacionarity přístupem autorů Hylleberg et al. se zatím ve standardních balíčcích statistických a ekonometrických systémů příliš nepoužívá.

Krok 1)

Po spuštění systému Gretl vybereme v hlavní nabídce tlačítko *„Soubor“* (v anglické verzi *„File“*), z nabídky zvolíme *„Soubory funkcí“* (v a.v. *„Function files“*), a klikneme na tlačítko *„Na serveru…“* (v a.v. *„On Server…“*), jak ukazuje obrázek 1.

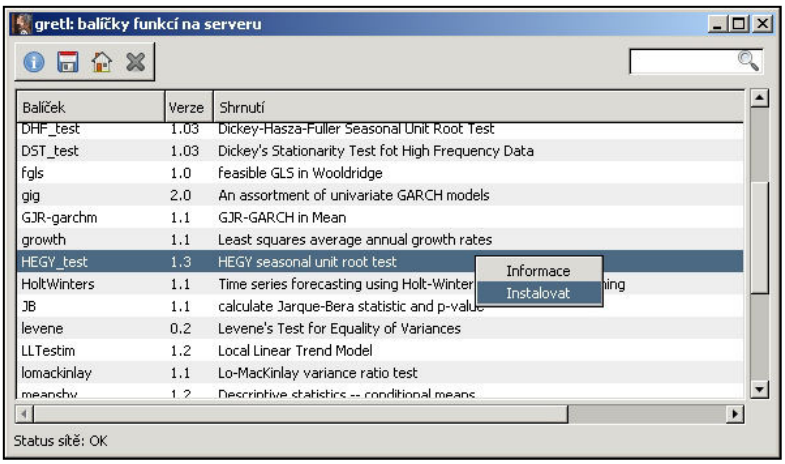


Obr. 1: Hlavní obrazovka systému Gretl

Krok 2)

Zobrazí se nové okno, které je zobrazeno na obrázku 2, ve kterém přibližně v polovině seznamu nalezneme položku pojmenovanou HEGY_test. Klikneme na ni pravým tlačítkem myši a vybereme nabídnutou možnost *„Instalovat“* (v a.v. *„Install“*). Po kliknutí na toto

tlačítko se již nic viditelného nestane. Okno můžeme zavřít a vrátíme se zpět k hlavní obrazovce systému Gretl.



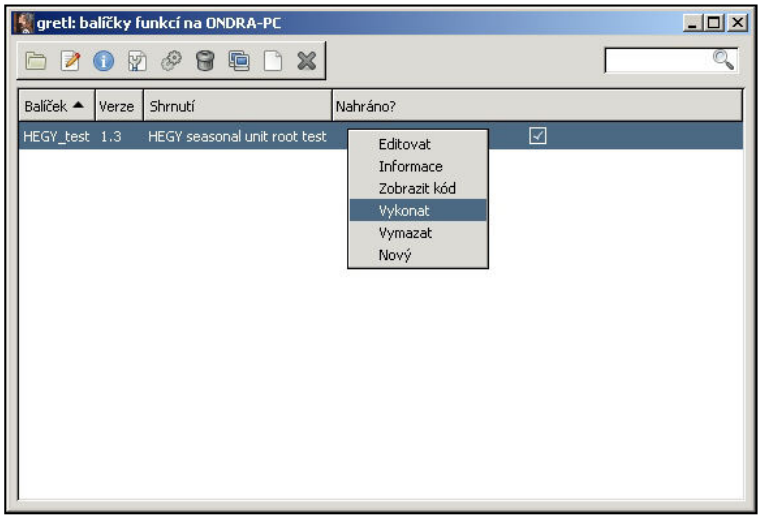
Obr. 2: Balíčky funkcí na serveru

Krok 3)

Po návratu do hlavní obrazovky Gretlu klikneme opět na tlačítko *„Soubor“* (v a.v. *„File“*), z nabídky zvolíme *„Soubory funkcí“* (v a.v. *„Function files“*), a klikneme na tlačítko *„Na místním počítači…“* (v a.v. *„On local machine…“*). Tlačítko je možno vidět i na obrázku 1.

Krok 4)

Zobrazí se nové okno, ve kterém bude seznam nainstalovaných doplňků na lokálním PC. Doplněk HEGY_test se spouští tak, že se na něj klikne pravým tlačítkem myši a vybere se možnost *„Vykonat“* (v a.v. *„Execute“*). Doplněk se spustí pouze v případě, že v základní obrazovce má uživatel otevřená data. Jinak zahlásí chybovou hlášku.



Obr. 3: Spuštění HEGY testu

3. Dosažené výsledky

Ze zvolených demografických časových řad, které byly ČSÚ publikovány s čtvrtletní frekvencí (počty sňatků, počty rozvodů, počty živě narozených osob, počty potratů celkem, počty zemřelých celkem, počty přistěhovalých a počty vystěhovalých), byla ve všech případech zamítnuta nulová hypotéza stacionarity na 5% hladině významnosti. Minimálně jeden z parametrů z_1, z_2, z_3 nebo z_4 byly na 5% hladině významnosti prokázány jako nulové. Výsledky shrnuje tabulka 1.

Tab. 1: Výsledky HEGY testu pro vybrané čtvrtletní demografické časové řady

SNATKY	**koeficient**	**směr. chyba**	**t-podíl**	**p-hodnota**
const	6345,35000	1841,35000	3,446	0,0010
time	-13,8541000	7,14161000	-1,940	0,0565
z1	-0,11011200	0,02968890	-3,709	0,0004
z2	-0,02753800	0,03003090	-0,917	0,3624
z3	-0,00762274	0,00969591	-0,786	0,4345
z4	-0,01323320	0,00961451	-1,376	0,1732
d4y_1	0,42566300	0,09966930	4,271	0,0001
ROZVODY				
const	5053,850	1419,7700	3,560	0,0007
time	-2,044170	4,2410800	-0,482	0,6314
z1	-0,160991	0,0455561	-3,534	0,0007
z2	-0,205903	0,0885868	-2,324	0,0231
z3	0,220502	0,0959667	2,298	0,0247
z4	0,325187	0,0923045	3,523	0,0008
d4y_1	0,190322	0,1228250	1,550	0,1259
ZIVE_NAROZENI				
const	2283,4700	952,52400	2,397	0,0193
time	10,920100	6,0440800	1,807	0,0752
z1	-0,0276901	0,0101182	-2,737	0,0079
z2	-0,1776990	0,0855798	-2,076	0,0416
z3	0,0333594	0,0391601	0,852	0,3973
z4	0,0416267	0,0388963	1,070	0,2883
d4y_1	0,6023090	0,0938876	6,415	0,0000
POTRATY_CELKEM				
const	2951,8100	710,68900	4,153	0,0001
time	-14,220900	4,6117300	-3,084	0,0030
z1	-0,0518291	0,0116865	-4,435	0,0000
z2	-0,1387010	0,0590383	-2,349	0,0217
z3	0,0190186	0,0888773	0,214	0,8312
z4	0,2785010	0,0824795	3,377	0,0012
d4y_1	0,3841340	0,0915528	4,196	0,0001
ZEMRELI_CELKEM				
const	18675,0000	7336,9800	2,545	0,0132
time	-22,428500	12,854700	-1,745	0,0855
z1	-0,1623660	0,0622092	-2,610	0,0111
z2	-0,2576140	0,0776817	-3,316	0,0015
z3	0,0212165	0,0507676	0,418	0,6773
z4	0,1215390	0,0496249	2,449	0,0169
d4y_1	-0,0316371	0,1187440	-0,266	0,7907
PRISTEHOVALI				
const	302,05700	825,39200	0,366	0,7155
time	17,259800	24,308600	0,710	0,4801
z1	-0,0302791	0,0187532	-1,615	0,1110
z2	-0,5104460	0,1236140	-4,129	0,0001
z3	0,2699530	0,1143450	2,361	0,0211
z4	0,4748740	0,1032190	4,601	0,0000
d4y_1	0,1024440	0,1251530	0,819	0,4159
VYSTEHOVALI				
const	188,09000	422,94100	0,445	0,6579
time	3,2107400	10,355200	0,310	0,7575
z1	-0,0257016	0,0176582	-1,456	0,1501
z2	-0,1948100	0,0829772	-2,348	0,0218
z3	0,6054780	0,1136490	5,328	0,0000
z4	0,3641730	0,1274180	2,858	0,0057
d4y_1	0,0978633	0,1227800	0,797	0,4282

Zdroj: vlastní konstrukce

4. Závěr

Výstupy ze softwaru jsou velmi přehledné a poskytují sofistikovaný důkaz o tom, že v demografické časové řadě s čtvrtletní frekvencí (která je navíc předpokládána s přítomnou sezonností) se vyskytuje trend. Po stažení doplňkové funkce HEGY se i po zavření software a ukončení všech dalších úkonů daný doplněk nevymaže. (Vymazání doplňku můžeme provést tak, že v nabídce *„Soubory funkcí“ – „Na místním počítači“* (v a.v. *„Function files“ – „On local machine“*), klikne na zvolený doplněk pravým tlačítkem myši a vybereme možnost *„Vymazat“* (v a.v. *„Delete“*), viz obrázek 3). Doplněk je však možné využít pouze pro řady s frekvencí čtvrtletní, nikoliv měsíční. To ovšem není zas až tak velký nedostatek, neboť s měsíční frekvencí se ve statistice příliš mnoho údajů nepublikuje.

Příspěvek byl zpracován v rámci projektu VŠE IGA 29/2011 „Analýza stárnutí obyvatelstva a dopad na trh práce a ekonomickou aktivitu“.

5. Literatura

[1] ARLT, J., ARLTOVÁ, M.: „*Ekonomické časové řady“*, Grada Publishing, 2007.

[2] Arlt, J., Arltová, M., Rublíková, E.: *„Analýza ekonomických časových řad s příklady“*, 2. vyd. Skripta VŠE Praha, 2004.

[3] BOX, G.E.P., JENKINS, G.: „*Time series analysis: Forecasting and control“*, San Francisco, Holden-Day, 1970.

[4] DICKEY, D.A., FULLER, W.A.: „Distribution of the Estimators for Autoregressive Time Series with a Unit Root“, *Journal of the American Statistical Association*, 74, 1979, str. 427–431.

[5] GRANGER, C.W.J.: „Some Properties of Time Series Data and Their Use in Econometric Model Specification“, *Journal of Econometrics* 16, 1981, str. 121-130.

[6] HYLLEBERG, S., ENGLE, R.F., GRANGER, C.W.J., YOO, B.S.: „Seasonal Integration and Cointegration“, *Journal of Econometrics* 44, 1990, str. 215-238.

[7] HARVEY, D.I., VAN DIJK, D.: „Sample size, lag order and critical values of seasonal unit root tests“, *Loughborough University – Institutional Repository*, 2003.
Zdroj: <https://dspace.lboro.ac.uk/dspace-jspui/bitstream/2134/352/3/erp03-09.pdf>
Publikováno: září 2003

Adresa autorů

Ondřej Šimpach, Ing.
VŠE v Praze, katedra demografie
Nám. W. Churchilla 4, 130 67 Praha 3
ondrej.simpach@vse.cz

Petra Dotlačilová, Ing.
VŠE v Praze, katedra demografie
Nám. W. Churchilla 4, 130 67 Praha 3
xdotp00@vse.cz

Jitka Langhamrová, doc., Ing., CSc.
VŠE v Praze, katedra demografie
Nám. W. Churchilla 4, 130 67 Praha 3
langhamj@vse.cz