

KRYSTYNA STRZAŁA

PANELOWE TESTY STACJONARNOŚCI – MOŻLIWOŚCI I OGRANICZENIA

1. WSTĘP

Coraz większa dostępność obszernych przekrojowo-czasowych baz danych zwróciła uwagę ekonometryków oraz ekonomistów zainteresowanych badaniami empirycznymi na wykorzystanie metod wnioskowania na podstawie danych przekrojowo-czasowych, obecnie popularnie zwanych panelowymi. Analogicznie jak w przypadku zastosowań klasycznych metod analizy ekonometrycznej, po okresie intensywnego wykorzystywania modeli panelowych przy jawnym lub ukrytym założeniu stacjonarności procesów podlegających analizie, ugruntowane zostało przekonanie, że dane panelowe nie są wolne od problemu braku stacjonarności szeregów czasowych indywidualnych jednostek występujących w panelu. Z drugiej strony wskazuje się na wnioskowanie o stacjonarności na podstawie badania szeregów zestawionych w postaci panelu danych jako na remedium udokumentowanej niskiej mocy jednowymiarowych testów pierwiastka jednostkowego (por. [21]). Ponieważ także panelowe testy nie są pozbawione ograniczeń i słabości, na co zwracamy uwagę w artykule, dla postawienia starannej diagnozy warto porównać wyniki badania stacjonarności z wykorzystaniem tak jednowymiarowych jak i panelowych testów pierwiastka jednostkowego.

Pierwsze testy panelowe autorstwa Levina i Lina, wzorowane na klasycznym już dzisiaj teście Dickeya-Fullera, powstały na początku lat 90., początkując rozwój metod wnioskowania w przypadku niestacjonarnych modeli panelowych. Bardzo szybko wnioskowanie o stacjonarności lub jej braku dla danych panelowych okazało się przydatną metodą empirycznej weryfikacji hipotez ekonomicznych, a w szczególności tzw. kontrowersji ekonomicznych, do których zaliczyć można hipotezę Parytetu Siły Nabywczej, dylemat Feldsteina i Horioki, czy też hipotezę konwergencji realnej krajów Unii Europejskiej.

2. CHARAKTERYSTYKA PANELOWYCH TESTÓW PIERWIASTKA JEDNOSTKOWEGO I STACJONARNOŚCI

Okres rozkwitu zastosowań jednowymiarowych testów stacjonarności, a w szczególności testu DF i ADF¹, przypada na lata 80. XX wieku, a pierwsze panelowe testy

¹ Ponieważ testy te zalicza się już obecnie do „klasyki” ekonometrii, nie będą w niniejszym artykule omawiane. Szczegółowe omówienie wraz z przykładami zastosowań można w literaturze polsko – języcznej znaleźć m.in. w podręcznikach i pracach: Charemza i Deadman [3], Osińska [25], Strzała [29], Syczewska [34], Welfe [36].

pierwiastka jednostkowego autorstwa Levina i Lina zostały opublikowane w 1992 r. Ogólną wersję procesu generującego dane panelowe, można zapisać jako:

$$y_{it} = \alpha_i + \delta_i t + \varphi_i y_{i,t-1} + \theta_i + \varepsilon_{i,t} \quad i = 1, 2, \dots, N, \quad t = 1, 2, \dots, T, \quad (1)$$

gdzie:
$$\varepsilon_{it} \sim \text{i.i.d.}(0, \sigma_\varepsilon^2). \quad (2)$$

Model (1)-(2) pozwala na uwzględnienie zróżnicowanych efektów indywidualnych (α_i), indywidualnie zróżnicowanego trendu liniowego (δ_i), heterogenicznego parametru autoregresyjnego (φ_i) oraz efektów czasowych (θ_i). W przypadku analiz stacjonarności, przedmiotem badania jest parametr (φ_i).

Na podstawie przeglądu literatury, do najczęściej stosowanych testów stacjonarności dla modeli panelowych można zaliczyć testy opracowane przez Levina i Lina [17], [18], Ima, Pesarana i Shina [14], [15], Harrisa i Tzavalisa [13], Maddali i Wu [21], Pedroniego z 1995 i 1997 roku oraz Hadri [12]. Pięć pierwszych testów zakłada brak stacjonarności w hipotezie zerowej, a w alternatywnej stacjonarność wszystkich części, lub też, co najmniej jednego szeregu, w zależności od testu. Ogólną postać hipotez można przedstawić następująco:

$$H_0: (\varphi_i - 1) = \rho_i = 0; \quad y_{it} \sim I(1), \quad (3)$$

$$H_A: (\varphi_i - 1) = \rho_i = 0; \quad y_{it} \sim I(0). \quad (4)$$

Levin i Lin [17] jako pierwsi przeprowadzili szeroko zakrojone badania nad właściwościami statystyk testów pierwiastka jednostkowego dla danych panelowych i zaproponowali testy dla szczegółowych wersji modelu (1)-(2)², wyróżniając sześć możliwości – od prostego homogenicznego modelu panelowego poprzez model z heterogenicznym przesunięciem³ do modelu zawierającego zarówno zróżnicowany efekt indywidualny jak i indywidualny współczynnik trendu, ale zawsze przy założeniu homogeniczności parametru autoregresyjnego. Przy czym warto podkreślić, że założenie homogeniczności parametru autoregresyjnego występuje tak w hipotezie zerowej jak i alternatywnej. Ponadto testy LL92 zakładają homoskedastyczność, brak autokorelacji i jednoczesnej korelacji składników losowych. Szczegółowe postacie regresji pomocniczych testów LL92 odpowiadających poszczególnym wersjom modelu (1)-(2) przedstawiają poniżej zamieszczone formuły (5)-(10):

Model 1:
$$\Delta y_{it} = \rho y_{i,t-1} + \varepsilon_{i,t}; \quad H_0: \rho = 0 \quad (5)$$

Model 2:
$$\Delta y_{it} = \rho y_{i,t-1} + \alpha + \varepsilon_{i,t}; \quad H_0: \rho = 0 \quad (6)$$

Model 3:
$$\Delta y_{it} = \rho y_{i,t-1} + \alpha + \delta t + \varepsilon_{i,t}; \quad H_0: \rho = 0, \delta = 0 \quad (7)$$

² Oznaczone w tekście jako LL92.

³ Inna stosowana nazwa to heterogeniczny dryf, co w efekcie oznacza zróżnicowany efekt indywidualny.

$$\text{Model 4:} \quad \Delta y_{it} = \rho y_{i,t-1} + \theta_i + \varepsilon_{i,t}; \quad H_0: \rho = 0 \quad (8)$$

$$\text{Model 5:} \quad \Delta y_{it} = \rho y_{i,t-1} + \alpha_i + \varepsilon_{i,t}; \quad H_0: \rho = 0, \alpha_i = 0 \text{ dla } \forall i \quad (9)$$

$$\text{Model 6:} \quad \Delta y_{it} = \rho y_{i,t-1} + \alpha_i + \delta_i t + \varepsilon_{i,t}; \quad H_0: \rho = 0, \delta_i = 0 \text{ dla } \forall i \quad (10)$$

Regresje pomocnicze w testach LL92 są szacowane przy zastosowaniu łącznej estymacji równań⁴ metodą MNK. Levin i Lin wykazali (por. [18]), że inaczej niż w przypadku testów pierwiastka jednostkowego dla indywidualnych szeregów czasowych, testy panelowe mają asymptotyczny rozkład normalny. Dla modeli (1)-(4) statystyki panelowego testu pierwiastka jednostkowego przy założeniu prawdziwości hipotezy zerowej mają rozkład $N(0, 1)$:

$$T\sqrt{N}\hat{\rho} \Rightarrow N(0, 2) \quad t_{\rho=0} \Rightarrow N(0, 1). \quad (11)$$

Dla modelu 5, podstawą wyznaczenia własności asymptotycznych statystyki jest przyjęcie założenia, że przy $N \Rightarrow \infty$ i $T \Rightarrow \infty$, szybkość zbieżania T do nieskończoności jest większa, tak że $\sqrt{N}/T \Rightarrow 0$. Asymptotyczne rozkłady statystyki testu przedstawiają formuły:

$$T\sqrt{N}\hat{\rho} + 3\sqrt{N} \Rightarrow N(0, 10, 2) \quad \sqrt{1.25}t_{\rho=0} + \sqrt{1.875N} \Rightarrow N\left(0, \frac{645}{112}\right) \quad (12)$$

W opracowaniu z 1993 r., Levin i Lin zaproponowali testy⁵ pierwiastka jednostkowego, w których uchylili założenie o homoskedastyczności i braku autokorelacji składnika losowego ε_{it} , pozostawiając założenie o braku jednoczesnej korelacji zakłóceń losowych. Analizując rozkłady asymptotyczne stwierdzili, że przy włączeniu odpowiedniej liczby opóźnień (analogicznie do testu ADF), statystyki zaproponowanych testów mają takie same rozkłady asymptotyczne jak testy bez rozszerzenia. W celu wyeliminowania zagregowanych efektów, na wstępnym etapie przeprowadza się transformację analizowanych szeregów do postaci odchyłeń od średnich grupowych, a następnie stosuje test ADF do indywidualnych szeregów przeprowadzając jednocześnie normalizację reszt⁶. Po oszacowaniu stosunku długookresowych do krótkookresowych odchyłeń standardowych dla każdego szeregu, a następnie wartości średniej tego ilorazu dla całego panelu, wyznacza się statystykę testu dla panelu danych (t_{ρ}^*), której rozkład po zastosowaniu odpowiedniej normalizacji jest asymptotycznie zbieżny do $N(0, 1)$. Wartości krytyczne oraz oszacowania współczynników stosowanych przy normalizacji są zawarte w publikacji Levina i Lina z 1993 roku.

Podstawową wadą testów LL92 oraz LL93 jest restrykcyjność hipotez, polegająca na założeniu homogeniczności parametru autoregresyjnego ρ tak w hipotezie zerowej jak i alternatywnej oraz na nie uwzględnieniu jednoczesnych korelacji. Warto zauważyć, że hipotezy zerowa oraz alternatywna w testach LL są następujące:

⁴ Ang. *pooled regression*.

⁵ Oznaczone jako LL93.

⁶ Testy LL93 są omówione szczegółowo w [31].

$$H_0: \rho_1 = \rho_2 = \dots = \rho_N = \rho = 0 \quad H_A: \rho_1 = \rho_2 = \dots = \rho_N = \rho < 0. \quad (13)$$

Dlatego też, testy te określa się jako typu „wszystko albo nic”. Podczas gdy sformułowanie hipotezy zerowej znajduje sensowne zastosowanie np. przy testowaniu konwergencji gospodarczej – implikując stwierdzenie, że żadna z badanych gospodarek nie podlega konwergencji, stwierdzenie implikowane przez hipotezę alternatywną jest zbyt restrykcyjne, gdyż zakłada, że wszystkie gospodarki podlegają procesowi konwergencji w tym samym tempie. W przypadku zastosowania testów LL do weryfikacji hipotezy Parytetu Siły Nabywczej⁷, poprzez badanie stacjonarności kursu wymiany, H_0 implikuje, że żaden z kursów nie spełnia założeń długookresowej PPP, a H_A – że tempo zbieżności do długookresowego PPP jest takie same dla wszystkich analizowanych kursów walut.

Z drugiej strony testy LL należą do najczęściej wykorzystywanych, głównie z tego względu, że po pierwsze były pierwszymi testami pierwiastka jednostkowego dla paneli danych, autorzy szczegółowo zbadali ich własności asymptotyczne, co zostało ostatecznie opublikowane przez Levina, Lina i Chu w *Journal of Econometrics* w 2002 roku, oraz łatwe w zastosowaniu, gdyż oprogramowane przez Chianga i Kao jako procedura NPT 1.2 (por. [4]) programu *GAUSS*, a także dostępne np. w pakiecie *STATA*.

Testami o bardzo podobnej konstrukcji są zaproponowane przez Quaha (por. [26])⁸ oraz Harrisa i Tzavalisa [13]⁹.

W teście zaproponowanym przez Harrisa i Tzavalisa, statystyka panelowego testu pierwiastka jednostkowego jest wyznaczona na podstawie znormalizowanego oszacowania parametru autoregresyjnego modelu typu (1), (5) i (6) Levina i Lina. Autorzy stosują trzy metody estymacji łącznej: LSP – *Least Squares Pooled*, LSDV – *Least Squares Dummy Variable* oraz LSDVT – *Least Squares Dummy Variable with Trend*. Przy zastosowaniu testów HT możliwe jest wnioskowanie o stacjonarności szeregów, tworzących panel danych na podstawie trzech wersji modelu generującego dane od najprostszego homogenicznego modelu panelowego (model 1 – LL) poprzez model zawierający heterogeniczny wyraz wolny (heterogeniczny efekt indywidualny) (model 5 – LL) do modelu zawierającego heterogeniczny wyraz wolny oraz indywidualnie zróżnicowany trend (model 6 – LL). Podstawową różnicą testów HT w stosunku do LL jest odmienna postać sprawdzianu testu¹⁰ oraz fakt opracowania własności asymptotycznych dla ustalonego T , przy założeniu, że $N \rightarrow \infty$.

Odmiernym w swojej konstrukcji jest test zaproponowany przez Ima, Pesarana i Shina ([14], [15])¹¹. Zasadnicza odmienność testu IPS w relacji do testów LL polega na rozluźnieniu założenia, że $\rho_1 = \rho_2 = \dots = \rho_N = \rho$ w hipotezie alternatywnej, czyli przyjęcie heterogenicznego parametru autoregresyjnego, co prowadzi do sformułowa-

⁷ Ang. *Purchasing Power Parity*, popularnie PPP.

⁸ Test Quaha nie będzie analizowany, gdyż bardzo rzadko jest wykorzystywany w badaniach empirycznych.

⁹ Por. [13], [2].

¹⁰ Sprawdzian testu HT jest wyznaczony jako znormalizowane oszacowanie parametru regresyjnego, podczas gdy statystyki proponowane przez LL są wyznaczane jako znormalizowany iloraz „t”.

¹¹ Test ten został po raz pierwszy opisany w *Discussion Paper* z 1997 roku, a następnie w *Journal of Econometrics* w 2003 r., oznaczony w dalszej części artykułu jako IPS.

nia regresji pomocniczej, analogicznej do modelu 5 w teście LL, lecz uwzględniającej heterogeniczny parametr autoregresyjny ρ_i . Regresja pomocnicza testu IPS przyjmuje postać:

$$\Delta y_{it} = \rho_i y_{i,t-1} + \alpha_i + \varepsilon_{i,t}; \quad H_0: \rho_i = 0 \text{ dla } \forall i. \quad (14)$$

Hipoteza alternatywna w teście IPS jest sformułowana jako:

$$H_A: \varphi_i < 0; \quad i = 1, 2, \dots, m \quad \varphi_i = 0 \quad i = m + 1, m + 2, \dots, N. \quad (15)$$

Odmienne do procedury LL, wykorzystującej podejście panelowe, czyli łączną estymację równań, test IPS polega na zastosowaniu testów pierwiastka jednostkowego do każdej z jednostek panelu osobno. Oznaczając indywidualne statystyki pierwiastka jednostkowego przez $t_{i,T}$, T – liczba obserwacji, $i = 1, 2, \dots, N$ liczba jednostek w panelu, a przez $t_{N,T}$ – statystykę panelową oraz przyjmując, że $E(t_{i,T}) = \mu$, $V(t_{i,T}) = \sigma^2$, możemy zapisać:

$$\Lambda_i = \sqrt{N} \frac{(t_{N,T} - \mu)}{\sigma} \Rightarrow N(0, 1), \text{ gdzie } t_{N,T} = \frac{1}{N} \sum_{i=1}^N t_{i,T}. \quad (16)$$

Oszacowania μ oraz σ wyznaczone na podstawie symulacji Monte Carlo są zawarte w tablicach (T. 3 i 4) publikacji Ima, Pesarana i Shina [14], [15]. W procedurze testu IPS zawarte jest założenie, że T – liczba obserwacji jest taka sama dla wszystkich jednostek panelu, a także, że $E(t_{i,T}) = \mu$, $V(t_{i,T}) = \sigma^2$ przyjmują taką samą wartość dla $\forall i$.

Ze względu na konstrukcję hipotezy alternatywnej, odrzucenie hipotezy zerowej na rzecz H_A nie oznacza, że w panelu nie występuje pierwiastek jednostkowy, lecz że hipoteza zerowa została odrzucona dla części jednostek, a dokładniej rzecz ujmując, dla co najmniej jednego indywidualnego szeregu czasowego w panelu danych.

Im, Pesaran i Shin zaproponowali również dwie wersje testu w przypadku autokorelacji składnika zakłócającego $\varepsilon_{i,t}$, analogiczne do opisanego powyżej, polegające na wyznaczeniu średniej indywidualnych testów ADF. W pierwszej z zaproponowanych wersji testu ($\bar{t}_{N,T}$) zakłada się występowanie takiej samej liczby rozszerzeń dla panelu danych. W drugiej wersji, istnieje możliwość zróżnicowania ilości rozszerzeń (p_i), a wyrażenia $E(t_{i,T})$ oraz $V(t_{i,T})$ zostają zastąpione przez odpowiednio obliczone średnie grupowe, których wartości wyznaczone na podstawie symulacji Monte Carlo, dla wybranych wartości (p_i) oznaczone jako $E(t_{i,T}, p_i)$ oraz $V(t_{i,T}, p_i)$ są zamieszczone w artykule Ima, Pesarana i Shina z 2003 roku. Powoduje to jednak nadal konieczność ograniczenia się do identycznej długości rozszerzeń w każdym z indywidualnych testów ADF¹² dla grupy badanych jednostek. Postać drugiej wersji testu IPS, czyli standaryzowana statystyka LR_T jest wyznaczona jako średnia wartość statystyk LR_i indywidualnych regresji pomocniczych typu ADF, czyli jako:

¹² Por. [21].

$$\Lambda_{LR} = \sqrt{N} \frac{(\bar{LR}_T - E(LR_T))}{\sqrt{Var(LR_T)}}, \quad (17)$$

gdzie $E(LR_T)$ oraz $Var(LR_T)$ ¹³ oznaczają wartości asymptotyczne – wartości oczekiwanej i wariancji średniej grupowej statystyki \bar{LR}_T . Przy prawdziwości hipotezy zerowej, obydwie statystyki $\bar{t}_{N,T}$ oraz \bar{LR}_T mają standaryzowany rozkład normalny dla $N \rightarrow \infty$ i $T \rightarrow \infty$ oraz przy założeniu, że $N/T \rightarrow 0$.

Do zalet testu IPS należy zaliczyć tak jego dostępność (oprogramowanie) jak i dostępność wartości krytycznych testu. Wady pierwszej z wersji wynikają z jej konstrukcji, gdyż statystyka $t_{N,T}$ testu IPS jest średnią arytmetyczną indywidualnych statystyk testu DF i/lub ADF poszczególnych jednostek panelu, a założenie, że liczba obserwacji (T) jest taka sama dla wszystkich jednostek panelu, a także, że $E(t_{i,T}) = \mu$, $V(t_{i,T}) = \sigma^2$ przyjmują taką samą wartość dla $\forall i$, powoduje, że można go stosować tylko do zbilansowanych paneli.

Maddala i Wu w artykule z 1999 r. zauważyli, że test IPS w swojej istocie polega na weryfikacji istotności wyników N niezależnych indywidualnych testów, zwracając jednocześnie uwagę, że na temat takiego podejścia, czyli wnioskowania na podstawie pewnej ilości niezależnych testów istnieje bardzo bogata literatura, datująca się już na lata 30. XX wieku. (por. [21], s. 636). Większość znanych procedur polega na odpowiednim łączeniu prawdopodobieństw empirycznych (ang. *p-values*). W przypadku, gdy statystyki testu są ciągłe, poziomy istotności π_i , ($i = 1, 2, \dots, N$) są niezależnymi zmiennymi, a wyrażenie $-2 \log_e \pi_i$ ma rozkład $\chi^2(2)$. Biorąc pod uwagę właściwość addytywności zmiennych o rozkładzie χ^2 , można zauważyć, że suma logarytmów prawdopodobieństw empirycznych $\gamma = -2 \sum_{i=1}^N \log_e \pi_i$ indywidualnych testów będzie miała rozkład $\chi^2(2N)$ – co stanowi istotę testu Fishera, znanego od 1932 r. W konsekwencji, Maddala i Wu zaproponowali stosowanie testu Fishera do badania występowania pierwiastków jednostkowych w panelu danych. W literaturze jest on często określany jako test Maddali i Wu (MW). Do zalet testu MW w porównaniu z IPS należy zaliczyć możliwość jego stosowania do niezbilansowanych paneli danych, a w przypadku stosowania ADF możliwość optymalizacji liczby wprowadzonych opóźnień (augmentacji), czego częściowo pozbawiony jest test IPS. Do wad – konieczność wyznaczenia indywidualnych wartości prawdopodobieństwa empirycznego na podstawie symulacji. Inną zaletą testu Fishera jest możliwość jego stosowania jako testu panelowego, w którym dla poszczególnych jednostek możemy zastosować dowolny ze znanych testów pierwiastka jednostkowego czy też stacjonarności (por. [21], s. 636).

Odmiennego rodzaju testem jest zaproponowany przez Hadri¹⁴ [12], w którym w hipotezie zerowej zakłada się stacjonarność, a w alternatywnej jej brak. Test H jest uogólnieniem testu KPSS [16] dla danych panelowych. Proces generujący dane może zawierać przesunięcie i/lub trend deterministyczny. Wyróżnia się trzy wersje testu, w zależności od przyjętych założeń, co do struktury stochastycznej procesu generującego dane panelowe: H-A – przy założeniu homogeniczności składników zakłócających

¹³ Wartości asymptotyczne są podane w artykule [15].

¹⁴ Oznaczany w dalszej części jako H.

w panelu, H-B – dopuszczający heterogeniczność¹⁵ składników zakłócających poszczególnych jednostek panelu; H-C – przy założeniu autokorelacji zakłóceń poszczególnych jednostek panelu. Proces generujący dane w teście H, można przedstawić następująco:

$$y_{it} = r_{it} + \varepsilon_{i,t} \quad (18A)$$

$$y_{it} = r_{it} + \beta_{it}t + \varepsilon_{i,t} \quad (18B)$$

$$y_{it} = r_{it-1} + \mu_{i,t} \quad (18C)$$

gdzie: y_{it} – analizowane szeregi czasowe, r_{it} – proces błędzenia losowego, ε_{it} i μ_{it} – składniki zakłócające o indywidualnych niezależnych rozkładach normalnych dla $i = 1, 2, \dots, N$ oraz $t = 1, 2, \dots, T$ z $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2 > 0, E(\varepsilon_{it}) = E(\mu_{it}) = 0, E(\mu_{it}^2) = \sigma_\mu^2 \geq 0$. Wartości początkowe r_{i0} są traktowane jako stałe i odgrywają rolę heterogenicznego efektu indywidualnego. Model (18B) zawiera oprócz efektu indywidualnego także zróżnicowany indywidualnie trend. Hipoteza zerowa, zakładająca stacjonarność jest postaci $H_0: \sigma_\mu^2 = 0$. Ponieważ z założenia $\varepsilon_{it} \sim \text{i.i.d.} \sim N(0, \sigma_\varepsilon^2)$, dla $i = 1, 2, \dots, N$, $t = 1, 2, \dots, T$, to przy prawdziwości hipotezy zerowej y_{it} jest procesem stacjonarnym wokół stałej (przesunięcia) w modelu (18A) oraz stacjonarnym wokół trendu w modelu (18B). Szczegółowe postacie hipotez są następujące:

$$H_0: \lambda = 0 \quad H_A: \lambda > 0 \quad \lambda = \frac{\sigma_\mu^2}{\sigma_\varepsilon^2}. \quad (19)$$

Statystyka stosowana do weryfikacji hipotez jest typu LM¹⁶ i jednocześnie LBI¹⁷ dla hipotezy $H_0: \lambda = 0$. Oznaczając reszty regresji pomocniczej (18A) oraz (18B) przez $\hat{\varepsilon}_{it}$, przy założeniu homogeniczności składników zakłócających w panelu danych, jednostronna statystyka typu LM jest określona jako:

$$LM = \frac{1}{N} \sum_{i=1}^n \left(\frac{\frac{1}{T^2} \sum_{t=1}^T S_{it}^2}{\hat{\sigma}_\varepsilon^2} \right) \quad (20)$$

gdzie S_{it} oznacza częściową sumę reszt: $S_{it} = \sum_{j=1}^t \hat{\varepsilon}_{ij}$, a $\hat{\sigma}_\varepsilon^2$ jest zgodnym estymatorem σ_ε^2 przy założeniu prawdziwości H_0 , na przykład postaci¹⁸:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}^2. \quad (21)$$

¹⁵ W nomenklaturze stosowanej przez Hadri, heterogeniczność dla $i = 1, 2, \dots, N$ oznacza, że wariancje indywidualnych zakłóceń mogą być zróżnicowane.

¹⁶ LM – skrót od ang. *Lagrange multiplier*.

¹⁷ LBI – skrót od ang. *locally best invariant*.

¹⁸ Dla prób skończonych proponuje się włączenie korekty ze względu na stopnie swobody.

Do wyprowadzenia właściwości asymptotycznych testu H, zakłada się zbieżność sekwencyjną: przy $T \rightarrow \infty - N \rightarrow \infty$. Założenie to jest uzasadnione dla paneli danych, w których wymiar czasowy jest znacznie większy od liczby jednostek w panelu ($T > N$).

Sprawdziany testu dla modeli (18A) i (18B), oznaczane jako Z_{jt} oraz Z_τ przy prawdziwości H_0 mają asymptotyczny rozkład normalny:

$$Z_{jt} = \frac{\sqrt{N}(LM_{jt} - \xi_{jt})}{\zeta_{jt}^2} \Rightarrow N(0,1) \quad Z_\tau = \frac{\sqrt{N}(LM_\tau - \xi_\tau)}{\zeta_\tau^2} \Rightarrow N(0,1) \quad (22)$$

gdzie ξ_{jt}, ζ_{jt}^2 oraz ξ_τ, ζ_τ^2 oznaczają wartość oczekiwaną i wariancję odpowiednio zdefiniowanych zmiennych losowych, a oszacowania ($\kappa_1 = \xi_{jt} = \frac{1}{6}, \kappa_2 = \xi_\tau = \frac{1}{45}, \kappa_1 = \xi_\tau = \frac{1}{15}, \kappa_2 = \zeta_\tau^2 = \frac{11}{6300}$) są podane w artykule Hadri [12].

Sprawdzian testu dla modelu (18A) oraz (18B) można łatwo rozszerzyć na proces generujący dane w przypadku, którego rozluźniamy założenie o homogeniczności zakłóceń. Wersja H-B testu Hadri wymaga wprowadzenia w zapisie (20) indywidualnych (heterogenicznych) wariancji składników zakłócających poszczególnych jednostek panelu, czyli wprowadzenia $\hat{\sigma}_{\varepsilon,j}^2$ w mianowniku formuły (20) w miejsce $\hat{\sigma}_\varepsilon^2$. Inną możliwością, prowadzącą do wersji H-C jest rozluźnienie założenia, że indywidualne składniki losowe są $\varepsilon_{it} \sim i.i.d.N(0, \sigma_\varepsilon^2)$, dla $\forall t$, czyli dopuszczenie możliwości, że składniki zakłócające podlegają autokorelacji, ale nie są jednocześnie skorelowane. Uwzględnienie autokorelacji indywidualnych składników zakłócających wymaga zastąpienia σ_ε^2 przez długookresową wariancję zdefiniowaną jako:

$$\sigma^2 = \frac{1}{N} \sum_{t=1}^N \lim_{T \rightarrow \infty} T^{-1} (S_{it}^2). \quad (23)$$

Jako zgodny estymator długookresowej wariancji można zastosować np. estymator zaproponowany przez Newey i Westa (por. [12], s. 155). Do zalet testu Hadri można zaliczyć fakt wyprowadzenia momentów rozkładu asymptotycznego w sposób analityczny, a nie na podstawie symulacji.

3. ROZMIAR I MOC PANELOWYCH TESTÓW PIERWIASKA JEDNOSTKOWEGO I STACJONARNOŚCI

Biorąc pod uwagę, że coraz częściej wskazuje się na wnioskowanie o stacjonarności na podstawie badania szeregów zestawionych w postaci panelu danych jako na remedium udokumentowanej niskiej mocy jednowymiarowych testów pierwiastka jednostkowego (por. [21], s. 631), bardzo ważnym zagadnieniem jest kwestia rozmiaru i mocy panelowych testów pierwiastka jednostkowego.

Większość autorów testów pierwiastka jednostkowego i stacjonarności publikuje jednocześnie wyniki analiz właściwości asymptotycznych rozkładów testów oraz wyniki analiz zachowania testów w skończonych próbach, czyli ich rozmiaru i mocy.

Harris i Tzavalis [13] na podstawie przeprowadzonych symulacji Monte Carlo (5000 replikacji) pokazują, że dla dużych wartości N w stosunku do T (lub też odpowiednio dużych wartości N) empiryczny rozmiar testu jest bardzo bliski wartości teoretycznej 0,05. Natomiast w przypadku małych wartości N i względnie dużych wartości T , rozkład empiryczny statystyki testu jest zbliżony do rozkładu zaprezentowanego przez Dickeya i Fullera. Natomiast dla małych wartości N i T następuje wyraźne pogorszenie właściwości testu i prawdopodobieństwo empiryczne wzrasta do 0,13 w przypadku zastosowania *LSP*, do 0,09 dla *LSDV* oraz do 0,06 dla *LSDVT*. Zgodnie z założeniami i przewidywaniami, moc testu wzrasta monotonicznie wraz ze wzrostem N i T . We wszystkich trzech rozpatrywanych przypadkach dla wartości $N > 100$ moc testu jest bliska wartości maksymalnej nawet dla bardzo małych wartości T . Moc testu poprawia się szybciej wraz ze wzrostem T dla danego N , niż wraz ze wzrostem N dla danego T . Autorzy zauważyli również, że moc testu osiąga swoją wartość maksymalną dla T znacznie mniejszego niż w przypadku indywidualnego testu DF, a jednocześnie wzrasta monotonicznie wraz ze wzrostem T i N . Ponadto wart odnotowania jest fakt, że przy założeniu, że T wzrasta asymptotycznie do nieskończoności, moc testu jest znacznie gorsza w małych próbach.

W przypadku testu Hadri rozmiar testu zależy tak od T jak i N . Właściwości testu dla skończonych prób autor (por. [12], s. 156-157) badał na podstawie symulacji Monte Carlo, stosując 20 tys. replikacji. Empiryczny rozmiar testu dla modelu (18A) jest bardzo bliski wartości teoretycznej 0,05 dla prób, w których wymiar czasowy $T > 10$. Na przykład dla $T = 15$ oraz $N = 15$, prawdopodobieństwo empiryczne odrzucenia hipotezy prawdziwej jest równe 0,0475. Wraz ze wzrostem T i N test staje się coraz bardziej precyzyjny w odniesieniu do rozmiaru, ale nie w sposób monotoniczny. Natomiast moc testu rośnie wraz ze wzrostem T lub N oraz dla wzrastających wartości T i N jednocześnie. W przypadku procesu generującego dane opisanego przez model (18B) test osiąga właściwy rozmiar dla $T > 25$. Dla $\lambda = (\sigma_{\mu}^2/\sigma_{\varepsilon}^2) > 0,01$ moc testu rośnie monotonicznie wraz ze wzrostem T lub N , oraz dla wzrastających wartości T i N jednocześnie.

Do jednego z pierwszych niezależnych porównań rozmiaru i mocy panelowych testów pierwiastka jednostkowego należy zaliczyć artykuł Maddala i Wu [21]. Autorzy porównują rozmiar i moc testów LL, IPS i zaproponowanego w tymże artykule zastosowania testu Fishera.

Porównania mocy testu LL i IPS dokonali również Im, Pesaran i Shin, ale Maddala i Wu podkreślają, że zaprezentowane przez nich porównanie nie jest wiarygodne, gdyż testy LL i IPS nie są bezpośrednio porównywalne pod względem mocy, gdyż hipotezy alternatywne w testach Levina i Lina oraz Ima, Pesarana i Shina się różnią. Maddala i Wu również porównują rozmiar i moc testów LL, IPS i Fishera, ale z zastrzeżeniem wzmiankowanym powyżej. Podstawowa różnica pomiędzy przeprowadzonymi przez Maddalę i Wu a Ima, Pesarana i Shina eksperymentami Monte Carlo (100 tys. replikacji), polega na założeniu przez MW możliwości występowania jednoczesnych korelacji składników zakłócających w procesie generującym obserwacje. Autorzy rozważają $N = 10, 25, 50, 100$ oraz $T = 10, 25, 50, 100$. Najważniejsze wnioski wynikające z przeprowadzonych symulacji oraz wyznaczenia empirycznych rozmiarów porównywanych testów są następujące:

– wyniki testów LL są najgorsze w zakresie rozmiaru testu¹⁹ – zgodnie z przewidywaniami, ale z zastrzeżeniem wzmiankowanym powyżej, że nie do końca właściwym jest porównywanie testów LL z IPS i Fishera ze względu na różnice w sformułowaniu hipotez alternatywnych;

– dla $T = 50$ i 100 oraz $N = 50$ i 100 test Fishera dominuje nad IPS w tym sensie, że charakteryzuje się mniejszymi odchyleniami od założonego poziomu istotności przy porównywalnej mocy testów tak dla DGP z przesunięciem jak i z przesunięciem i trendem.

Autorzy zwracają jednocześnie uwagę, że niedoszacowanie liczby opóźnień (rzędu $AR(p)$) w regresji pomocniczej testu ADF ma istotny wpływ na rozmiar testów, potwierdzając jednocześnie wcześniejsze spostrzeżenia Ima, Pesarana i Shina. Przeszacowanie rzędu autoregresji łagodzi problem odchylenia rozmiaru dla wszystkich analizowanych testów (LL, IPS, Fishera) jednakże kosztem zmniejszenia ich mocy.

W kolejnym eksperymencie Maddala i Wu porównują rozmiar i moc testów LL, IPS i Fishera dla panelu danych ($N = 25$, $T = 25$) zawierającego stacjonarne i niestacjonarne szeregi, zwiększając liczbę szeregów stacjonarnych od 1 do 2, 4, 8, 10 i 12. Wnioski są następujące:

- test Fishera ma największą moc we wszystkich rozpatrywanych przypadkach,
- relatywna przewaga testu Fishera rośnie wraz ze wzrostem liczby szeregów stacjonarnych,
- przy braku jednoczesnych korelacji składników zakłócających moc IPS jest większa od testu Fishera,
- wystąpienie heteroskedastyczności i autokorelacji składników zakłócających nie powoduje poważnych zakłóceń rozmiaru i mocy testów,
- w przypadku wystąpienia jednoczesnych korelacji składników zakłócających (między obserwacjami i/lub jednostkami w panelu) występują poważne zaburzenia rozmiaru i mocy wszystkich rozpatrywanych testów, przy czym dla dużego T w porównaniu z N odchylenie rozmiaru testu Fishera jest stosunkowo małe.

Innym istotnym opracowaniem poruszającym zagadnienia rozmiaru i mocy panelowych testów pierwiastka jednostkowego jest artykuł Straussa i Yigita z 2003 r. {[28]}, koncentrujący się na wskazaniu konsekwencji występowania jednoczesnych korelacji składników zakłócających procesu generującego obserwacje. W takim przypadku Im, Pesaran i Shin zalecają przekształcenie danych do postaci odchyłeń od średniej grupowej. Autorzy na podstawie przeprowadzonych symulacji Monte Carlo (5000) dla jednocześnie skorelowanych N ($N = 10, 25, 50, 100$) szeregów czasowych ($T = 50$) wyznaczyli odchylenie rozmiaru testu IPS oraz dostosowane do empirycznego rozmiaru wartości krytyczne dla jednakowych (homogenicznych) oraz zróżnicowanych (heterogenicznych) współczynników korelacji. Przeprowadzili porównanie mocy testów dla heterogenicznych współczynników korelacji. Wyniki przeprowadzonych symulacji wskazują, że zarówno zwiększenie skorelowania jak niejednorodności współczynników korelacji powoduje zwiększenie odchyłeń empirycznego rozmiaru testu od jego teore-

¹⁹ Autorzy zwracają uwagę na to, że porównanie mocy testów nie są uprawnione, jako że występują duże różnice w zakresie rozmiaru, a nie dokonano wyznaczenia mocy testu dostosowanej do jego empirycznego rozmiaru.

tycznego poziomu oraz znaczące podwyższenie wartości krytycznych (co do modułu). Co więcej Autorzy stwierdzili, że odchylenia rozmiaru rosną wraz ze wzrostem liczby jednostek w panelu (N). Prowadzi to do stwierdzenia, że przekształcenie danych do postaci odchyleń od średniej grupowej, zalecane przez Ima, Pesarana i Shina nie eliminuje odchyleń rozmiaru testu, jeżeli mamy do czynienia z heterogenicznymi korelacjami procesu generującego obserwacje panelowe, co może w konsekwencji prowadzić do niewłaściwego wnioskowania w badaniach empirycznych. Warto przy tym przytoczyć za Straussem i Yigitem, że bardzo często wykorzystywane w badaniach empirycznych zbiory danych: konwergencja – Maddison, PPP – dane IFS, czy też zbiór BLS charakteryzują się niejednorodnymi współczynnikami korelacji o zmienności pomiędzy 0,55 a 0,65. Na podstawie przeprowadzonych badań Autorzy stwierdzają, że w przypadku danych Maddisona oraz IFS, zastosowanie testu IPS do danych w postaci odchyleń od średnich grupowych pozwala na odrzucenie hipotezy zerowej na poziomie istotności $\alpha = 0,01$, podczas gdy zastosowanie wartości krytycznych wyznaczonych przy założeniu niejednorodnych korelacji nie pozwala na jej odrzucenie nawet na poziomie $\alpha = 0,10$.

A co jeszcze ważniejsze i bardziej niebezpieczne, z badań Straussa i Yigita wynika, że w takim przypadku moc testu IPS nie rośnie wraz ze wzrostem \sqrt{N} , a większe panele charakteryzują się większym odchyleniem empirycznego rozmiaru testu od poziomu teoretycznego.

4. PRZYKŁADY WNIOSKOWANIA Z WYKORZYSTANIEM PANELOWYCH TESTÓW STACJONARNOŚCI

Jednym z pierwszych i nadal najczęstszych obszarów wykorzystania panelowych testów pierwiastka jednostkowego i stacjonarności jest weryfikacja hipotezy Parytetu Siły Nabywczej, którą można określić jako jedną z najstarszych i nadal wzbudzających kontrowersje hipotez ekonomicznych. W swojej najbardziej znanej postaci pochodzi od sformułowania Cassela z 1916 r. Kontrowersje wokół PPP biorą się m.in. z faktu, że badania empiryczne z wykorzystaniem całego dostępnego arsenału podejść i technik ekonometrycznych ostatnich 40 lat dostarczają niejednoznacznych wyników²⁰. Innym, cieszącym się dużym zainteresowaniem obszarem badań ostatnich lat jest weryfikacja jednej z nowszych kontrowersyjnych hipotez, która doczekała się już określenia „łami-główka”, „paradoks”, czy też dylemat Feldsteina – Horioki.

W przypadku weryfikacji hipotezy PPP, wczesne zastosowania testów integracji typu ADF, analiz kointegracji relacji PPP, jak też procedury Johansena w większości wskazywały na konieczność odrzucenia długookresowego PPP, szczególnie dla okresu płynnych kursów walutowych. Jednym z problemów podnoszonych już na tym etapie była niska moc testów pierwiastka jednostkowego. Jako remedium już w drugiej połowie lat 90. XX wieku, część badaczy zaczęła stosować wczesne panelowe testy pierwiastka jednostkowego autorstwa Levina i Lina [18] do weryfikacji hipotezy PPP, uzyskując najczęściej potwierdzenie jej prawdziwości. Do prac charakteryzujących

²⁰ Najszerszy przegląd wyników empirycznych oraz postaci hipotezy PPP znaleźć można w publikacji [27], w języku polskim m.in. w [25], [30], [34].

ten nurt wykorzystania panelowych testów pierwiastka jednostkowego należy zaliczyć artykuły i opracowania MacDonalda [20], Oha [24], czy też Wu [37]. Ujmując chronologicznie, następną z publikacji jest artykuł autorstwa Coakleya i Fuertes z 1997 r., [5] w którym autorzy omawiają wyniki weryfikacji hipotezy PPP dla panelu złożonego z 11 krajów, wykorzystując test IPS do miesięcznych szeregów czasowych realnego kursu wymiany, zdefiniowanego jako:

$$\pi_t = s_t - p_t + p_t^* \quad (24)$$

gdzie: s_t oznacza logarytm nominalnego kursu wymiany, wyrażonego jako ilość jednostek waluty krajowej na 1 US\$, a p_t i p_t^* oznaczają odpowiednio logarytmy wybranej miary poziomu cen krajowych i zagranicznych (USA). W literaturze przedmiotu stosuje się również alternatywną interpretację równania (22) w postaci odchyleń od PPP²¹, co można przedstawić jako:

$$s_t = \alpha + \beta_1 p_t + \beta_2 p_t^* + \mu_t \quad (25)$$

gdzie μ_t oznacza składnik biało-szumowy.

Często stosowaną procedurą badawczą jest wykorzystanie podejścia Engle'a i Grangera, czyli np. zastosowanie testu ADF, czy też Phillipsa-Perrona do zbadania stacjonarności reszt relacji (25). Można również zastosować procedurę Johansena do zbadania, czy występuje kointegracja s_t , p_t , p_t^* , w przypadku badania silnej wersji PPP uzupełniona o warunek zakładający jednoczesne występowanie symetrii i proporcjonalności, czyli $\beta_1 = -\beta_2 = 1$ ²².

Weryfikację hipotezy PPP, wspomniani autorzy – Coakley i Fuertes przeprowadzili na podstawie procedury Johansena, oraz procedury Engle'a i Grangera przy wykorzystaniu szeregu klasycznych indywidualnych testów pierwiastka jednostkowego oraz panelowego testu zaproponowanego przez Ima, Pesarana i Shina. Dla panelu krajów ($i = 1, 2, \dots, N$) równanie pomocnicze testu ADF bez trendu dla realnego kursu wymiany, można zapisać jako:

$$\Delta s_{it} = \alpha_i + \beta_i s_{i,t-1} + \sum_{j=1}^{p_i} \gamma_{ij} \Delta s_{i,t-j} + v_{it} \quad (t = 1, 2, \dots, T). \quad (26)$$

Baza danych obejmowała 11 krajów (G10+Szwajcarię), a miesięczne szeregi czasowe nominalnego kursu wymiany oraz wskaźników cen WPI i CPI dotyczyły okresu lipiec 1973 do czerwca 1996, co stanowiło łącznie 276 obserwacji dla każdej ze zmiennych. Zgodnie z przewidywaniami, zastosowanie indywidualnego testu ADF nie pozwoliło na odrzucenie hipotezy zerowej o występowaniu pierwiastka jednostkowego dla wszystkich procesów generujących dane z wyjątkiem WPI dla UK na 5% poziomie istotności. Zastosowanie testu ADF jak też Phillipsa-Perrona zgodnie

²¹ Omówienie typowych relacji stosowanych do weryfikacji PPP, znaleźć można m.in. w [30], [34].

²² Szczegółowe omówienie postaci relacji służących weryfikacji silnej oraz słabej wersji PPP, zostały zamieszczone m.in. w [30].

z procedurą Engle'a i Grangera również wskazywało na brak stacjonarności relacji PPP dla wszystkich jednostek badanego panelu. Wykorzystanie procedury Johansena dało również mieszane wyniki, statystyka wartości własnych wskazywała na jeden wektor kointegrujący w 7 przypadkach, ale jednocześnie należało odrzucić warunek symetryczności i proporcjonalności w 5 z 7 analizowanych pod tym kątem przypadków. Przechodząc do zastosowania testu IPS, zgodnie z sugestią Ima, Pesarana i Shina, dokonano przekształcenia danych do postaci odchyleń od średnich grupowych, w celu wyeliminowania zagregowanych efektów. Zastosowanie statystyki $\bar{t}_{N,T}$ z uwzględnieniem dostosowania do ewentualnych korelacji między krajami, pozwoliło na odrzucenie hipotezy zerowej dla 10-ciu realnych kursów wymiany. Także zastosowanie statystyki $L\bar{R}_T$ pozwoliło Autorom na odrzucenie hipotezy zerowej. Uzyskując zgodne wyniki dla obydwu statystyk Autorzy postawili wniosek, że realne kursy wymiany charakteryzują się stacjonarnością, zwłaszcza, że prezentowane wyniki są również zgodne z wynikami MacDonalda (por. [20]) oraz Oha (por. [24]) dla próby kończącej się w latach 1992/93.

Dylemat Feldsteina i Horioki należy natomiast do jednej z najnowszych kontrowersji ekonomicznych, gdyż pojawił się wraz z opublikowaniem przez M. Feldsteina i C. Horiokę artykułu w *The Economic Journal* w 1980 roku. Artykuł ten wywołał szeroko publikowaną na łamach czasopiśmiennictwa ekonomicznego dyskusję na temat znaczenia mobilności kapitału, sposobu pomiaru oraz obserwowanych w tym zakresie barier i ograniczeń.

Przyjmując powszechnie wyznawaną opinię, że kapitał jest nieskończenie mobilny, czyli, że każdy kraj może pożyczać lub też udzielać pożyczek innym krajom przy obowiązującej światowej realnej stopie procentowej (r), Feldstein i Horioka postawili hipotezę, że oszczędności oraz inwestycje krajowe nie są skorelowane.

Do weryfikacji hipotezy zaproponowali wykorzystanie przekrojowej regresji stopy inwestycji i oszczędności, postaci:

$$i_i = \alpha + \beta s_i + u_i \quad (1)$$

gdzie małe litery oznaczają udziały inwestycji (I_i) oraz oszczędności (S_i) w PKB w kraju i . Rozpatrując oszczędności oraz inwestycje jako wielkości brutto a także netto, dla próby obejmującej 16 krajów OECD w latach 1960-1974, FH uzyskali statystycznie istotne oszacowanie tzw. wskaźnika zatrzymania oszczędności²³, $b = 0,89$, które zinterpretowali jako świadczące o niskiej mobilności kapitału.

Artykuł Feldsteina-Horioki wywołał bardzo szeroki oddźwięk. Kolejno po sobie następujące badania empiryczne z lat 1980-1991, wykorzystujące do weryfikacji dylematu regresję przekrojową potwierdzały wnioski Feldsteina i Horioki. Ten okres zainteresowania dylematem FH jest wyczerpująco opisany w artykule Obstfelda [22] oraz Tesar [35]. W latach 80. zainteresowanie badaczy zwróciło się w kierunku wykorzystania analizy szeregów czasowych, nie dostarczając także rozwiązania dylematu. Charakterystyczną cechą tego okresu badań jest występowanie znacznego zróżnicowania oszacowań współczynnika zatrzymania oszczędności, którego oceny kształtują się od

²³ Ang. *savings – retention coefficient*.

0,13 do 0,92 (por. [22]), czy też od 0,025 dla Luksemburga do 1,182 dla Szwajcarii, a generalnie wskazują na znacząco wyższe oszacowania współczynnika β dla krajów wysoko rozwiniętych w porównaniu z rozwijającymi się (por. [7]). Jako reprezentatywne dla tego nurtu, a jednocześnie zawierające krytyczny przegląd wyników empirycznych można wymienić opracowanie Obstfeldta i Rogoffa z 2000 roku.

W latach 80. XX wieku część badaczy zaczęła stosować do weryfikacji dylematu podejście panelowe. Do najczęściej cytowanych prac należących do tego nurtu zalicza się artykuł Feldsteina z 1983 roku oraz opracowania [1] i [11].

W latach 90. XX wieku zaczęto coraz częściej zwracać uwagę na problem braku stacjonarności stopy inwestycji i oszczędności. Akceptując obowiązywanie warunku „płynności rachunku bieżącego”, należy uznać, że proces generujący obserwacje rachunku bieżącego²⁴ bilansu płatniczego jest procesem zintegrowanym $I(0)$, a w związku z tym oszczędności i inwestycje, które są niestacjonarne muszą być skointegrowane z wektorem kointegrującym równym $[1, -1]$ (por. [6], [33]). Zastosowanie konwencjonalnych metod badania stacjonarności oraz kointegracji przyniosło zróżnicowane wyniki, w zależności od zastosowanej metody oraz kraju poddanego analizie. W tym przypadku tak jak w wielu innych, jako jedną z możliwych interpretacji zróżnicowania wyników wskazuje się na niską moc klasycznych testów stacjonarności i poszukuje rozwiązań poprzez zastosowanie panelowych testów pierwiastka jednostkowego do zweryfikowania stacjonarności CA.

Jako reprezentatywny przykład badania stacjonarności stopy inwestycji, oszczędności oraz udziału CA w PKB przy wykorzystaniu jednowymiarowych oraz panelowych testów pierwiastka jednostkowego można przytoczyć wyniki zamieszczone w artykule Coakleya, Kulasi i Smitha z 1996 roku [6]. Badaniu zostały poddane szeregi dla 23 krajów OECD za okres 1960-1992. Przeciętne, statystycznie istotne oszacowanie współczynnika zatrzymania oszczędności wyznaczone na podstawie regresji przekrojowej wyniosło 0,731, średnia z regresji wykorzystujących szeregi czasowe 0,719. Przy zastosowaniu testu ADF do indywidualnych szeregów autorzy odrzucili hipotezę zerową o występowaniu pierwiastka jednostkowego na 5% poziomie istotności dla stopy oszczędności brutto – dla 1 kraju, stopy inwestycji brutto – 4 krajów oraz dla 10 w przypadku CA. Warto podkreślić, że chociaż odrzucenie hipotezy o braku stacjonarności CA występowało wielokrotnie częściej niż dla stopy inwestycji czy oszczędności, tylko dla mniej niż połowy krajów można było uznać, że CA jest stacjonarny. Przechodząc do podejścia panelowego autorzy zastosowali statystykę $\bar{t}_{N,T}$ testu IPS, której wartości przy zróżnicowaniu specyfikacji (z przesunięciem i/lub trendem) pozwoliły w każdym przypadku na odrzucenie hipotezy zerowej o braku stacjonarności udziału CA w PKB (por. [6], s. 625). Tym samym wyniki badań Coakleya, Kulasi i Smitha przyczyniły się do upowszechnienia interpretacji, że wysokie oszacowania współczynnika zatrzymania oszczędności otrzymywane niezależnie od zastosowanej techniki wynikają z „ograniczenia budżetowego”, czyli warunku płynności rachunku bieżącego, a nie stanowią adekwatnej miary mobilności kapitału przynajmniej w przypadku krajów wysoko uprzemysłowionych.

²⁴ Ang. *current account balance*, oznaczany w dalszej części jako CA.

Nieco odmiennie kształtują się oszacowania tak współczynnika zatrzymania oszczędności jak też charakterystyki szeregów czasowych w przypadku krajów mniej uprzemysłowionych. Badania własne²⁵ autorki artykułu obejmujące 15 + 8 nowo-przyjętych członków Unii Europejskiej za okres 1960-2001²⁶, a w przypadku analiz rachunku bieżącego 22 kraje²⁷, przy zastosowaniu indywidualnego badania stacjonarności wykazały, że tylko w przypadku stopy inwestycji dla Szwecji, nie można było odrzucić hipotezy zerowej o występowaniu pierwiastka jednostkowego w indywidualnych procesach generujących dane. Dla salda rachunku bieżącego nie przeprowadzono badania indywidualnych szeregów dla nowych krajów członkowskich ze względu na zbyt krótkie szeregi czasowe. Wnioskowanie przeprowadzone z wykorzystaniem trzech testów panelowych, testu IPS, LLC oraz H pozwoliło na sformułowanie następujących obserwacji.

Wyniki testu H w każdej z jego wersji tak dla stopy inwestycji, oszczędności oraz udziału salda rachunku bieżącego w PKB dla poziomów zmiennych wskazywały na odrzucenie hipotezy zerowej, zakładającej stacjonarność – na każdym z rutynowo stosowanych poziomów istotności (0,01; 0,05; 0,1). Zgodność wskazań pary testów (IPS oraz H) uzyskano w przypadku stopy inwestycji tak dla panelu złożonego z 23 jak i 15 krajów, co pozwoliło na sformułowanie wniosku, że włączenie nowych 8 państw do UE, nie zmieniło charakteru niestacjonarności całego panelu danych – występuje panelowy pierwiastek jednostkowy. Wyniki badania stacjonarności dla pierwszych różnic, potwierdzają stopień zintegrowania $I(1)$ stopy inwestycji dla 23 krajów, ale są niejednoznaczne dla 15 „starych” krajów UE²⁸.

W przypadku stopy oszczędności oraz salda na rachunku bieżącym również należy sformułować wniosek o braku stacjonarności procesów generujących obserwacje tak dla „15” jak i dla rozszerzonej UE, gdyż pomimo tego, że wskazania testów dla poziomów są wzajemnie sprzeczne, wyniki testów dla pierwszych różnic wskazują na brak stacjonarności stopy oszczędności oraz udziału salda na rachunku bieżącym w PKB dla „23” krajów. Test Hadri, przy założeniu heterogeniczności zakłóceń jednostek panelu wskazał na zintegrowanie rzędu 1.

Odmienność uzyskanych wyników w stosunku do Coakleya i innych [6] wynika przede wszystkim z faktu rozpatrywania innego okresu oraz innego składu grupy krajów, gdyż warto zauważyć, że wzmiankowani autorzy rozpatrywali najwyżej rozwinięte kraje świata (OECD) w innym okresie czasowym, gdyż w latach 1960-1992.

5. PODSUMOWANIE

Przedstawione panelowe testy pierwiastka jednostkowego i stacjonarności cieszą się dużym zainteresowaniem środowiska naukowego, co pozwala przypuszczać, że

²⁵ Por. [31], [32].

²⁶ Najdłuższe z szeregów obejmują lata 1960-2001.

²⁷ Wyłączenie Luksemburga wynika z bardzo krótkiego szeregu czasowego CA oraz faktu, że ze względu na wyjątkowo rozbudowany sektor finansowy w tym kraju, należy go sklasyfikować jako „obserwację nietypową”.

²⁸ W tym przypadku wyniki testów LLC oraz IPS pozwalają odrzucić hipotezę, że szereg jest $I(2)$ na rzecz $I(1)$, a jednocześnie test H na 1% poziomie istotności nie pozwala odrzucić hipotezy, że $y_t \sim I(1)$.

w niedługim czasie powstaną ich udoskonalone wersje. Zaprezentowane wyniki badań dotyczące weryfikacji hipotezy Parytetu Siły Nabywczej z wykorzystaniem testów panelowych wskazują jednocześnie na ich efektywność oraz skuteczność. Własne wyniki badania stacjonarności szeregów obserwacji stopy inwestycji oraz oszczędności dla „starej” jak i „poszerzonej” UE, z zastosowaniem testów jednowymiarowych jak i panelowych wskazują na brak stacjonarności badanych procesów, czego nie można traktować jako sprzeczne z wynikami Coakleya, Kulasi i Smitha, ze względu na różnice rozpatrywanego okresu czasu oraz odmienny skład grupy badanych krajów.

Uniwersytet Gdański

LITERATURA

- [1] AmirKhalkhali S.A.A. Dar, [1993], *Testing for capital mobility: a random coefficient approach*, „Empirical Economics”, Vol. 3, s. 523-541.
- [2] Baltagi B.H., [2001], *Econometric Analysis of Panel Data*, John Wiley&Sons, Chichester.
- [3] Charemza W.W., Deadman D.F., [1997], *New Directions in Econometric Practice*, Edwar Elgar, Cheltenham, Lyme.
- [4] Chiang M-H., C. Kao, [2000], *Nonstationary Panel Time Series using NPT 1.2 – A User Guide*, Center for Policy Research, Syracuse University.
- [5] Coakley J., Fuertes A.M., [1997], *New panel unit root tests of PPP*, „Economic Letters”, Vol. 57, s. 17-22.
- [6] Coakley J., Kulasi F., Smith R., [1996], *Current account solvency and the Feldstein-Horioka puzzle*, „Economic Journal”, Vol. 106, s. 620-627.
- [7] Coakley J., Kulasi F., Smith R., [1998], *The Feldstein Horioka puzzle and capital mobility: a review*, „International Journal of Finance and Economics”, Vol. 3, s. 169-188.
- [8] Feldstein M., [1983], *Domestic Saving and International capital Movements in the Long Run and the Short Run*, „European Economic Rreview”, Vol. 21, s. 129-151.
- [9] Feldstein M., Horioka C., [1980], *Domestic saving and international capital flows*, „The Economic Journal”, Vol. 90, s. 314-329.
- [10] Fleissig A.R., Strauss J., [2000], *Panel unit root tests of purchasing power parity for price indices*, „Journal of International Money and Finance”, Vol. 19, s. 489-506.
- [11] Ghosh A., [1995], *International capital mobility amongst major industrialised countries: Too little or too much*, „The Economic Journal”, Vol. 105, s. 107-128.
- [12] Hadri K., [2000], *Testing for stationarity in heterogenous panel data*, „The Econometrics Journal”, Vol. 3, s. 148-161.
- [13] Harris R.D.F., Tzavalis E., [1999], *Inference for unit roots in dynamic panels where the time dimension is fixed*, „Journal of Econometrics”, Vol. 91, s. 201-226.
- [14] Im K., Pesaran H., Shin Y., [1997], *Testing for Unit Roots in Heterogenous Panels*, mimeo, Department of Applied Economics, University of Cambridge.
- [15] Im K., Pesaran H., Shin Y., [2003], *Testing for unit roots in heterogenous panels*, „Journal of Econometrics”, Vol. 115, s. 53-74.
- [16] Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y., [1992], *Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?*, „Journal of Econometrics”, Vol. 54, s. 159-178.
- [17] Levin A., Lin Ch-F., [1992], *Unit Root Test in Panel Data: Asymptotic and Finite Sample Properties*, Discussion Paper No. 93, University of California at San Diego.
- [18] Levin A., Lin Ch-F., [1993], *Unit Root Tests in Panel Data: New Results*, Discussion Paper, No. 56, University of California at San Diego.
- [19] Levin A., Lin Ch-F., Chu Ch-Sh. J., [2002], *Unit root tests in panel data: asymptotic and finite sample properties*, „Journal of Econometrics”, Vol. 108, s. 1-24.

- [20] MacDonald R., [1995], *Long-Run Exchange Rate Modelling: A Survey of the Recent Evidence*, International Monetary Fund Staff Papers, 42, s. 437-489.
- [21] Maddala G.S., Wu S., [1999], *A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test*, „Oxford Bulletin of Economics and Statistics”, Special Issue, s. 631-652.
- [22] Obstfeldt M., [1986], *Capital mobility in the world economy: Theory and measurement*, „Carnegie-Rochester Conference Series on Public Policy”, Vol. 24, s. 55-104.
- [23] Obstfeldt M., Rogoff K., [2000], *The six major puzzles in international macroeconomics: is there a common cause?*, NBER working paper 7777.
- [24] Oh K.-Y., [1996], *Purchasing Power parity and unit root tests using panel data*, „Journal of International Money and Finance”, Vol. 15, s. 405-418.
- [25] Osińska M., [2000], *Ekonometryczne modelowanie oczekiwań gospodarczych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- [26] Quah D., [1994], *Exploiting Cross-Section Variation for Unit Root Inference in Dynamic Data*, „Economics Letters”, Vol. 44, s. 9-19.
- [27] Rogoff K., [1996], *The purchasing power parity puzzle*, „Journal of Economic Literature”, Vol. 34, s. 647-668.
- [28] Strauss J., Yigit T., [2003], *Shortfalls of panel unit root testing*, „Economics Letters”, Vol. 81, s. 309-313.
- [29] Strzała K., [1994], *Nieklasyczne metody sterowania optymalnego*, Wydawnictwo Uniwersytetu Gdańskiego, Sopot.
- [30] Strzała K., [2002], *Weryfikacja hipotez makroekonomicznych – ewolucja podejść na przykładzie PPP*, [w:] T. Kufel, M. Piłatowska (red.), *Analiza szeregów czasowych na początku XXI wieku*, s. 103-114, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- [31] Strzała K., [2005a], *Relacja inwestycji i oszczędności w krajach Unii Europejskiej – weryfikacja empiryczna z wykorzystaniem podejścia panelowego*, s. 141-156, Materiały i Prace Wydziału Zarządzania Uniwersytetu Gdańskiego, Nr 1.
- [32] Strzała K., [2005b], *Relacja inwestycji, oszczędności i salda rachunku bieżącego w krajach Unii Europejskiej – weryfikacja empiryczna z zastosowaniem podejścia panelowego*, s. 25-36, [w:] T. Kufel, M. Piłatowska (red.), *Dynamiczne modele ekonometryczne*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- [33] Strzała K., [2006], *Dylemat Feldsteina – Horioki -- teoria i empiria*, s. 251-259, [w:] P. Chrzan (red.), *Metody Matematyczne, Ekonometryczne i Informatyczne w Finansach i Ubezpieczeniach*, Wydawnictwo Akademii Ekonomicznej im. K. Adamieckiego, Katowice.
- [34] Syczewska E.M., [2007], *Ekonometryczne modele kursów walutowych*, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie.
- [35] Tesar L., [1991], *Saving, investment and international capital flows*, „Journal of International Economics”, Vol. 31, s. 55-78.
- [36] Welfe A., [2003], *Ekonometria. Metody i ich zastosowanie*, wyd. III, PWE, Warszawa.
- [37] Wu Y., [1996], *Are Real Exchange Rates Stationary? Evidence from a Panel Data Test*, „Journal of Money, Credit and Banking”, Vol. 28, s. 54-63.

Praca wpłynęła do redakcji w grudniu 2008 r.

PANELOWE TESTY STACJONARNOŚCI – MOŻLIWOŚCI I OGRANICZENIA

Streszczenie

Celem artykułu jest przedstawienie oraz omówienie podobieństw i różnic w konstrukcji panelowych testów stacjonarności i pierwiastka jednostkowego oraz wskazanie znanych z literatury możliwości ich wykorzystania do interpretacji hipotez ekonomicznych. Wskazane zostaną ich zalety a także ograniczenia. Przedmiotem analiz porównawczych będzie pięć panelowych testów stacjonarności, tj. Levina, Lina (LL, 1992, 1993), Ima, Pesarana i Shina (IPS, 1997), Harrisa i Tzavalisa (HT,1999), Maddali i Wu (MW, 1999)

oraz Hadri (H, 2000). Rozważania teoretyczne zostaną zilustrowane przykładami wykorzystania panelowych testów pierwiastka jednostkowego i stacjonarności do weryfikacji hipotezy Parytetu Siły Nabywczej oraz dylematu Feldsteina i Horioki.

Słowa kluczowe: panelowe testy integracji, homogeniczne i heterogeniczne dane panelowe, hipoteza PPP, dylemat Feldsteina – Horioki.

PANEL UNIT ROOT TESTS – POTENTIAL AND LIMITATIONS

Summary

The paper aims at presenting panel unit root tests and pointing out the main similarities and differences among them. Shortfalls and potentials of PURT are discussed in a paper. Theoretical side is illustrated with the recent applications of PURT for evaluation and interpretation of economic hypothesis i.e. PPP and Feldstein-Horioka dilemma. Five PURT are considered, namely Levin, Lin (LL, 1992, 1993), Im, Pesaran and Shin (IPS, 1997), Harris and Tzavalis (HT,1999), Maddala and Wu (MW, 1999) and Hadri (H, 2000).

Key words: Panel unit root tests, homogenous and heterogenous panel data, PPP hypotheses, Feldstein-Horioka dilemma