Dynamiczne Modele Ekonometryczne
X Ogólnopolskie Seminarium Naukowe, 4–6 września 2007 w Toruniu
Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu

Barbara Pawełek
Akademia Ekonomiczna w Krakowie

Normalizacja zmiennych a dopuszczalność prognoz zmiennej syntetycznej

1. Wprowadzenie

W badaniach porównawczych dotyczących złożonych zjawisk ekonomicznych często wykorzystuje się zmienne syntetyczne. Jednym z elementów dynamicznej analizy jest budowa ścieżek rozwoju badanego zjawiska w oparciu o funkcje trendu.

Głównym celem pracy jest przedstawienie wyników badań nad wpływem normalizacji zmiennych diagnostycznych, ze stałymi parametrami, na wartości ocen *ex ante* błędów prognoz punktowych i przedziałowych, obliczanych na podstawie niektórych funkcji trendu dla wybranych zmiennych syntetycznych. Prezentowane badania były poświęcone rozszerzeniu rozważań przedstawionych w pracach Pawełek (2004), (2006). Dotyczyły one funkcji trendu liniowego dla addytywnej zmiennej syntetycznej. Rozszerzenie polegało na włączeniu do analizy trendów nieliniowych oraz rozważeniu multiplikatywnej zmiennej syntetycznej. Obie zmienne syntetyczne zarówno addytywna, jak i multiplikatywna były typu bezwzorcowego.

W rozważaniach przyjęto, że finalny zbiór zawiera m $(m > 1)$ zmiennych diagnostycznych X_{jt}, o realizacjach x_{ijt} $(i = 1, ..., n;\ j = 1, ..., m;\ t = 1, ..., k)$, opisujących złożone zjawisko ekonomiczne. Założono, że zmienne diagnostyczne są stymulantami mierzonymi na silnych skalach. W celu zapewnienia porównywalności w czasie transformowanych zmiennych normalizację przeprowadzono ze stałymi parametrami, tzn.:

$$y_{ijt} = \frac{x_{ijt} - a_j}{b_j}, \left(i = 1, ..., n;\ j = 1, ..., m;\ t = 1, ..., k\right), \tag{1}$$

gdzie: x_{ijt} – wartość wejściowej zmiennej X_{jt} dla obiektu O_i w okresie t; y_{ijt} – wartość znormalizowanej zmiennej Y_{jt}; a_j i b_j – wartości parametrów normalizacyjnych dla zmiennej X_{jt}.

Bezwzorcowa zmienna syntetyczna Z_t będąca średnią znormalizowanych zmiennych Y_{jt} przyjmuje następujące wartości:

– w wersji addytywnej (z przekształceniem normalizacyjnym liniowym, tzn. $a_j \neq 0$ dla $j = 1, ..., m$ lub ilorazowym, tzn. $a_j = 0$ dla $j = 1, ..., m$):

$$z_{it} = \frac{1}{m}\sum_{j=1}^{m} y_{ijt} = \frac{1}{m}\sum_{j=1}^{m} \frac{x_{ijt} - a_j}{b_j} \quad \left(i = 1, ..., n;\ t = 1, ..., k\right), \tag{2}$$

– w wersji multiplikatywnej (z przekształceniem ilorazowym):

$$z_{it} = \left(\prod_{j=1}^{m} y_{ijt}\right)^{\frac{1}{m}} = \left(\prod_{j=1}^{m} \frac{x_{ijt}}{b_j}\right)^{\frac{1}{m}} \quad \left(i = 1, ..., n;\ t = 1,...,k\right). \tag{3}$$

Rozważono cztery funkcje trendu, tzn.: liniowy i logarytmiczny dla addytywnej zmiennej syntetycznej oraz potęgowy i wykładniczy dla multiplikatywnej zmiennej syntetycznej. Przyjęto założenie, że zmienne X_{jt} charakteryzują się podobnymi przebiegami w czasie. To znaczy, że funkcje trendu najlepiej dopasowane do rzeczywistych danych zmiennych X_{jt} dla obiektu O_i są tego samego typu.

W tab. 1–2 zamieszczono wzory na prognozy punktowe i przedziałowe oraz mierniki *ex ante* dla trendów: liniowego, logarytmicznego, potęgowego i wykładniczego. Zaprezentowane równania ukazują związki między wyborem formuły normalizacyjnej i wynikami prognozowania zmiennej syntetycznej.
Należy pamiętać, że obliczanie ocen *ex ante* względnego błędu punktowej prognozy (9) oraz względnej precyzji przedziałowej prognozy (12) jest dopuszczalne tylko dla zmiennych mierzonych na skali ilorazowej. Mierniki te znajdują zatem zastosowanie tylko w przypadku, gdy wszystkie zmienne diagnostyczne są mierzone na skali ilorazowej i dokonano normalizacji za pomocą ilorazowego przekształcenia, tzn. $a_j = 0$ $\left(j = 1, ..., m\right)$.

2. Wpływ parametrów normalizacyjnych na wartości mierników *ex ante* – przypadek addytywnej zmiennej syntetycznej

Zakładając tę samą formułę normalizacyjną i budowę addytywnej zmiennej syntetycznej, można zauważyć duże podobieństwo między wynikami uzyskanymi dla trendów liniowego i logarytmicznego (por. tab. 1).

Tabela 1. Wzory dla funkcji trendu liniowego lub logarytmicznego oszacowanej dla bezwzorcowej addytywnej zmiennej syntetycznej Z_t

Zmienna Z_t	Równanie $(i = 1, ..., n;\ j = 1, ..., m;\ t = 1, ..., k;\ T = k + 1, ...)$	Nr
Punktowa prognoza	$z_{iT}^{P} = \frac{1}{m}\sum_{j=1}^{m}\frac{x_{ijT}^{P} - a_j}{b_j}$	(4)
Ocena *ex ante* średniego błędu	– dla trendu liniowego: $s_{D_T i}(Z_T) = \sqrt{\frac{1}{m^2}\sum_{j=1}^{m}\frac{s_{D_T i}^{2}(X_{jT})}{(b_j)^2} + \frac{2w_T}{m^2}\frac{k}{k-2}\sum_{j=s+1}^{m}\sum_{s=1}^{m-1}\frac{1}{b_j b_s}\mathrm{cov}(e_{ijt}, e_{ist})}$	(5)
	gdzie: $w_T = 1 + \frac{1}{k} + \frac{(T-\bar{t})^2}{\sum_{t=1}^{k}(t-\bar{t})^2}$	(6)
	– dla trendu logarytmicznego: $s_{D_T i}(Z_T) = \sqrt{\frac{1}{m^2}\sum_{j=1}^{m}\frac{s_{D_T i}^{2}(X_{jT})}{(b_j)^2} + \frac{2w'_T}{m^2}\frac{k}{k-2}\sum_{j=s+1}^{m}\sum_{s=1}^{m-1}\frac{1}{b_j b_s}\mathrm{cov}(e_{ijt}, e_{ist})}$	(7)
	gdzie: $w'_T = 1 + \frac{1}{k} + \frac{(\ln T - \overline{\ln t})^2}{\sum_{t=1}^{k}(\ln t - \overline{\ln t})^2}$	(8)
Ocena *ex ante* względnego błędu	$\hat{V}_{D_T i}(Z_T) = \frac{s_{D_T i}(Z_T)}{z_{iT}^{P}}\quad (z_{iT}^{P} \neq 0)$	(9)
Przedziałowa prognoza	$P\left(z_{iT}^{P} - t_{\alpha,r}\cdot s_{D_T i}(Z_T) < z_{iT} < z_{iT}^{P} + t_{\alpha,r}\cdot s_{D_T i}(Z_T)\right) = \gamma_{iT}$	(10)
Precyzja	$d_{I_{iT}^{P}} = t_{\alpha,r}\cdot s_{D_T i}(Z_T)$	(11)
Względna precyzja	$V_{I_{iT}^{P}} = \frac{t_{\alpha,r}\cdot s_{D_T i}(Z_T)}{z_{iT}^{P}}\quad (z_{iT}^{P} \neq 0)$	(12)

Symbole we wzorach oznaczają, odpowiednio, e_{ijt} i e_{ist} – reszty w funkcjach trendu, odpowiednio, liniowego lub logarytmicznego zmiennych X_{jt} i X_{st} $(j \neq s)$; x_{ijT}^{P} – punktowa prognoza zmiennej X_{jT}; $s_{D_T i}^{2}(X_{jT})$ – ocena *ex ante* wariancji błędu punktowej prognozy x_{ijT}^{P}; γ_{iT} - wiarygodność przedziałowej prognozy; $t_{\alpha,r}$ – wartość krytyczna w rozkładzie *t* Studenta.

Źródło: obliczenia własne.

Po podstawieniu do wzoru (9) równań (4) i (5) – dla trendu liniowego lub (7) – dla trendu logarytmicznego okazuje się, że znając wartości: k, m, w_T – dla trendu liniowego lub w'_T – dla trendu logarytmicznego, $s_{e_{ij}}^{2}$, $\mathrm{cov}(e_{ijt}, e_{ist})$, x_{ijT}^{P} oraz a_j i b_j można sprawdzić, jaką wartością oceny *ex ante* względnego błędu będzie charakteryzowała się punktowa prognoza z_{iT}^{P} (4) uzyskana w wyniku ekstrapolacji funkcji trendu liniowego lub logarytmicznego oszacowanej dla wartości addytywnej zmiennej syntetycznej. Analizując wzory (5) i (7) można stwierdzić, że im większe wartości parametru skalującego b_j $(j=1,...,m)$, tym mniejsza wartość spodziewanego błędu $s_{D_T i}(Z_T)$ punktowej prognozy addytywnej zmiennej syntetycznej. Pozostałe elementy tych wzorów, w przypadku oszacowania trendów, odpowiednio, liniowych lub logarytmicznych dla zmiennych diagnostycznych, są znane i nie ulegają zmianom na kolejnych etapach badania. W przypadku oceny *ex ante* względnego błędu prognozy, nie można sformułować podobnego wniosku, gdyż w mianowniku wzoru (9) znajduje się wartość punktowej prognozy (4), która zależy od poziomu parametru skalującego. Miernikami *ex ante* rzędu dokładności przedziałowej predykcji są: wiarygodność predykcji, precyzja predykcji oraz względna precyzja predykcji. Przy ustalonej wiarygodności predykcji γ_{iT} wnioskowanie jest tym dokładniejsze, im krótszy jest przedział prognozy (tzn. wartość miernika precyzji $d_{I_{iT}^{P}}$ (11) jest mniejsza). Podobnie, jak w przypadku punktowej predykcji, znając elementy występujące we wzorach na punktową prognozę (4), ocenę *ex ante* średniego błędu tej prognozy (wzór (5) lub (7)) oraz zadając satysfakcjonujący poziom wiarygodności przedziałowej prognozy można sprawdzić precyzję i względną precyzję możliwych do wyznaczenia przedziałowych prognoz addytywnej zmiennej syntetycznej. Przy zadanym prawdopodobieństwie γ_{iT}, zwiększanie wartości parametru skalującego b_j $(j=1,...,m)$, powoduje zmniejszanie wartości miernika precyzji (11). Oznacza to zwiększanie dokładności wnioskowania w przyszłość na podstawie przedziałowych prognoz. Niestety podobnie, jak w przypadku punktowej predykcji, nie oznacza to zwiększania precyzji w ujęciu względnym (12).

3. Wpływ parametrów normalizacyjnych na wartości mierników *ex ante* – przypadek multiplikatywnej zmiennej syntetycznej

W prognozowaniu zjawisk ekonomicznych wykorzystuje się także trendy nieliniowe, w tym trend potęgowy lub wykładniczy. Podobnie, jak dla trendów liniowego i logarytmicznego, zostało pokazane, jak poprzez wybór parametrów normalizacyjnych można wpływać na dopuszczalność prognoz. W przypadku multiplikatywnych trendów przyjęto, że zmienna syntetyczna ma także multiplikatywną postać (3). Założono także, że zmienne diagnostyczne w celu sprowadzenia ich do porównywalności, zostały poddane ilorazowemu przekształceniu.

Tabela 2. Wzory dla funkcji trendu potęgowego lub wykładniczego oszacowanej dla bezwzorcowej multiplikatywnej zmiennej syntetycznej Z_t

Zmienna Z_t	Równanie $(i = 1, ..., n;\ j = 1, ..., m;\ t = 1, ..., k;\ T = k + 1, ...)$	Nr
Punktowa prognoza	$z_{iT}^P = \left(\prod_{j=1}^{m} \frac{x_{ijT}^P}{b_j} \right)^{\frac{1}{m}}$	(13)
Ocena *ex ante* średniego błędu	– dla trendu potęgowego: $s_{D_T i}(Z_T) \approx z_{iT}^P \sqrt{\frac{1}{m^2} \sum_{j=1}^{m} s_{D'_T i}^2 (X_{jT}) + \frac{2w'_T}{m^2} \frac{k}{k-2} \sum_{j=s+1}^{m} \sum_{s=1}^{m-1} \mathrm{cov}(e'_{ijt}, e'_{ist})}$	(14)
	– dla trendu wykładniczego: $s_{D_T i}(Z_T) \approx z_{iT}^P \sqrt{\frac{1}{m^2} \sum_{j=1}^{m} s_{D'_T i}^2 (X_{jT}) + \frac{2w_T}{m^2} \frac{k}{k-2} \sum_{j=s+1}^{m} \sum_{s=1}^{m-1} \mathrm{cov}(e'_{ijt}, e'_{ist})}$	(15)
Ocena *ex ante* względnego błędu $\left(z_{iT}^P \neq 0\right)$	– dla trendu potęgowego: $\hat{V}_{D_T i}(Z_T) = \sqrt{\frac{1}{m^2} \sum_{j=1}^{m} s_{D'_T i}^2 (X_{jT}) + \frac{2w'_T}{m^2} \frac{k}{k-2} \sum_{j=s+1}^{m} \sum_{s=1}^{m-1} \mathrm{cov}(e'_{ijt}, e'_{ist})}$	(16)
	– dla trendu wykładniczego: $\hat{V}_{D_T i}(Z_T) = \sqrt{\frac{1}{m^2} \sum_{j=1}^{m} s_{D'_T i}^2 (X_{jT}) + \frac{2w_T}{m^2} \frac{k}{k-2} \sum_{j=s+1}^{m} \sum_{s=1}^{m-1} \mathrm{cov}(e'_{ijt}, e'_{ist})}$	(17)
Przedziałowa prognoza	por. (10)	
Precyzja	por. (11)	
Względna precyzja $\left(z_{iT}^P \neq 0\right)$	– dla trendu potęgowego: $V_{I_{iT}^P} = t_{\alpha,r} \sqrt{\frac{1}{m^2} \sum_{j=1}^{m} s_{D'_T i}^2 (X_{jT}) + \frac{2w'_T}{m^2} \frac{k}{k-2} \sum_{j=s+1}^{m} \sum_{s=1}^{m-1} \mathrm{cov}(e'_{ijt}, e'_{ist})}$	(18)
	– dla trendu wykładniczego: $V_{I_{iT}^P} = t_{\alpha,r} \sqrt{\frac{1}{m^2} \sum_{j=1}^{m} s_{D'_T i}^2 (X_{jT}) + \frac{2w'_T}{m^2} \frac{k}{k-2} \sum_{j=s+1}^{m} \sum_{s=1}^{m-1} \mathrm{cov}(e'_{ijt}, e'_{ist})}$	(19)

Podobnie, jak w tab. 1 przy czym: e'_{ijt} i e'_{ist} – reszty w addytywnych modelach transformowanych zmiennych X_{jt} i X_{st} $(j \neq s)$; $s_{D'_T i}^2(X_{jT})$ – ocena *ex ante* wariancji błędu logarytmu punktowej prognozy x_{ijT}^P.

Źródło: obliczenia własne.

Na podstawie szczegółowych wyników otrzymanych w trakcie badań zauważono, że wariancja reszt w addytywnej funkcji, będącej transformacją trendu potęgowego lub wykładniczego opisującego multiplikatywną zmienną syntetyczną zbudowaną w oparciu o znormalizowane ilorazowym przekształceniem wejściowe dane, nie zależy od parametru skalującego b_j $(j = 1, ..., m)$ przyjętego na etapie normalizacji. Także ocena *ex ante* średniego błędu prognozy otrzymanej z addytywnej funkcji, w rozważanym przypadku, nie zależy od przyjętego parametru normalizacyjnego b_j $(j = 1, ..., m)$.

Po uwzględnieniu wzoru na prognozę (13) można zauważyć, że ocena *ex ante* średniego błędu prognozy multiplikatywnej zmiennej syntetycznej (wzór (14) lub (15)) zależy od parametru normalizacyjnego b_j $(j = 1, ..., m)$ (por. tab. 2). Po podstawieniu do wzoru na ocenę *ex ante* względnego błędu (9) równań (13) i (14) – dla trendu potęgowego lub (15) – dla trendu wykładniczego okazuje się, że znając wartości: k, m, $w_T^{'}$ – dla trendu potęgowego lub w_T – dla trendu wykładniczego, $s_{e_{ij}^{'}}^2$ oraz $\operatorname{cov}\left(e_{ijt}^{'}, e_{ist}^{'}\right)$ można sprawdzić, jaką wartością oceny *ex ante* względnego błędu będzie charakteryzowała się punktowa prognoza z_{iT}^P (13) uzyskana w wyniku ekstrapolacji trendu potęgowego lub wykładniczego oszacowanego dla wartości zmiennej syntetycznej.

Przedziałowe prognozowanie wraz z oceną jego dokładności przebiega tak samo, jak w przypadku trendów liniowego i logarytmicznego.

Wzory na ocenę *ex ante* względnych błędów (9) i (12) warto zapisać w postaci uwzględniającej równania (14) lub (15). Wzory (16) i (18) – dla trendu potęgowego oraz (17) i (19) – dla trendu wykładniczego pokazują, że spodziewany względny błąd punktowych prognoz i przewidywana względna precyzja przedziałowych prognoz, w analizowanym przypadku, nie zależą od parametru skalującego przyjętego w ilorazowym przekształceniu normalizacyjnym.

4. Przykład na umownych danych

Rozważaniom poddano trend potęgowy. Założono, że zmienne diagnostyczne są mierzone na skali ilorazowej. Dane zostały poddanie ilorazowemu przekształceniu normalizacyjnemu. Zmienna syntetyczna została zbudowana zgodnie z multiplikatywną formułą (3). Umowne realizacje zmiennych diagnostycznych zapisano w tab. 3.

Tabela 3. Wartości zmiennych diagnostycznych (umowne dane)

Zmienna diagnostyczna	Rok							
	2000	2001	2002	2003	2004	2005	2006	2007
X_{1t}	50	63	68	77	80	87	90	92
X_{2t}	30	50	63	81	91	106	116	127
X_{3t}	80	90	93	99	103	104	106	108

Źródło: opracowanie własne.

Rolę parametru skalującego w ilorazowym przekształceniu pełnił rozstęp (por. tab. 4). Korzystając ze wzorów (13)–(19) obliczono prognozy oraz bezwzględne i względne mierniki *ex ante* dokładności prognoz (por. tab. 5–6). Otrzymane wyniki informują, że prognozy punktowe i przedziałowe multiplikatywnej zmiennej syntetycznej obliczone na podstawie potęgowego trendu będą tworzyły dopuszczalną ścieżkę rozwoju badanego zjawiska.

Tabela 4. Wartości miary opisowej pełniącej rolę parametru skalującego

Przypadek	Miara opisowa	Zmienna diagnostyczna		
		X_{1t}	X_{2t}	X_{3t}
P.1	$b_j = R_{X_{jt}}\ (j=1,2,3)$	42	97	28
P.2	$b_j = R_{X_{jt}}\ (j=1,2,3)$	100	150	60

Źródło: obliczenia własne.

Tabela 5. Wartości punktowych prognoz oraz ocen *ex ante* średnich i względnych błędów prognoz dla zmiennych diagnostycznych i syntetycznej

Prognoza oraz ocena *ex ante* średniego i względnego błędu	Rok	Zmienna diagnostyczna			Zmienna syntetyczna	
		X_{1T}	X_{2T}	X_{3T}	Z_T	
					P.1	P.2
x_{jT}^P lub z_T^P	$T=9$	96.474	138.608	110.487	2.348	1.180
	$T=10$	99.533	149.102	112.178	2.444	1.228
	$T=11$	102.384	159.278	113.729	2.533	1.273
$s_{D_T}(X_{jT})$ lub $s_{D_T}(Z_T)$	$T=9$	1.971	3.161	1.381	0.039	0.020
	$T=10$	2.076	3.470	1.431	0.042	0.021
	$T=11$	2.178	3.781	1.480	0.044	0.022
$\hat{V}_{D_T}(X_{jT})$ lub $\hat{V}_{D_T}(Z_T)$	$T=9$	0.020	0.023	0.013	0.017	0.017
	$T=10$	0.021	0.023	0.013	0.017	0.017
	$T=11$	0.021	0.024	0.013	0.017	0.017

Źródło: obliczenia własne.

Tabela 6. Przedziałowe prognozy (dla $\gamma_T = 0{,}95$) oraz wartości mierników precyzji i względnej precyzji przedziałowych prognoz zmiennej syntetycznej

Rok	$\left(z_T^P - t_{\alpha,r} \cdot s_{D_T}(Z_T),\ z_T^P + t_{\alpha,r} \cdot s_{D_T}(Z_T)\right)$		$d_{I_T^P}$		$V_{I_T^P}$	
	P.1	P.2	P.1	P.2	P.1	P.2
$T=9$	(2.252; 2.445)	(1.131; 1.228)	0.096	0.048	0.041	0.041
$T=10$	(2.342; 2.546)	(1.176; 1.279)	0.102	0.051	0.042	0.042
$T=11$	(2.425; 2.641)	(1.218; 1.327)	0.108	0.054	0.043	0.043

Źródło: obliczenia własne.

W celu prezentacji skutków zmiany parametru skalującego zwiększono wartości parametru $b_j\ (j=1,2,3)$ (por. przypadek P.2 w tab. 4). Wartości prognoz zmiennej syntetycznej i oceny *ex ante* średniego błędu uległy zmniejszeniu, natomiast oceny *ex ante* względnego błędu pozostały bez zmian (por. tab. 5). Zmalały wartości miernika precyzji przedziałowych prognoz, poprawiając tym samym ich dokładność wnioskowania o przyszłym rozwoju badanego zjawiska (por. tab. 6). Natomiast względna precyzja nie zmieniła się w stosunku do poziomu obliczonego w przypadku P.1.

5. Podsumowanie

W pracy pokazano, że – dla niektórych funkcji trendu zmiennej syntetycznej – im większe wartości parametru skalującego, tym m.in. mniejsza wartość spodziewanego średniego błędu punktowej prognozy oraz przy zadanej wiarygodności, większa dokładność (precyzja) wnioskowania w przyszłość na podstawie prognozy przedziałowej.

Można sformułować wniosek, iż decyzja dotycząca sposobu normalizacji zmiennych, w tym wybór parametrów normalizacyjnych, ma znaczący wpływ na to, czy prognozy obliczone na podstawie funkcji trendu oszacowanej dla zmiennej syntetycznej będą dopuszczalne ze względu na wartości mierników ex ante dokładności prognoz.

Przedstawione rozważania skłaniają do prowadzenia dalszych badań nad wpływem wyboru formuły normalizacyjnej na wyniki syntetycznych badań porównawczych w ujęciu dynamicznym.

Wykazany wpływ normalizacji zmiennych na dopuszczalność prognoz zmiennej syntetycznej, gdy kryterium jest wartość oceny *ex ante* błędu, może być kolejnym głosem w dyskusji toczącej się wokół podejmowanych prób odpowiedzi na pytanie: jak dobierać parametry normalizacyjne, aby z jednej strony spełnione były formalne wymogi przy konstrukcji zmiennych syntetycznych, a z drugiej – zachowana interpretacja ekonomiczna tych zmiennych.

Literatura

Pawełek, B. (2004), Wpływ normalizacji zmiennych diagnostycznych na dopuszczalność prognoz zmiennej syntetycznej, *Przegląd Statystyczny*, z. 4, 81–87.

Pawełek, B. (2006), Spostrzeżenia dotyczące wpływu normalizacji zmiennych na ocenę ex ante błędu prognozy, *Zeszyty Naukowe AE w Krakowie*, nr 726.