

STEFAN SCHMIDT I JAN STECZKOWSKI
Wyższa Szkoła Rolnicza
K r a k ó w

OBLICZANIE KORELACJI MIĘDZY SZEREGAMI DYNAMICZNYMI

(Uwagi polemiczne)

W specjalnym dodatku do „Zagadnień Ekonomiki Rolnej”, numer 4/1962, któremu dano tytuł: „Produkcja rolna i rynek w latach 1956—1961”, na stronach 81 do 83, poddana została m. in. krytyce metoda obliczania korelacji użyta przez S. Schmidta i J. Steczkowskiego w opracowaniu pt. „Produkcja żywca i mięsa wieprzowego w świetle statystyki” (Zagadnienia Ekonomiki Rolnej nr 4/1960).

Pragniemy ustosunkować się do tych uwag krytycznych z zastrzeżeniem, że pozostawiamy na uboczu zagadnienie merytoryczne dotyczące antagonizmu między produkcją hodowlaną bydła i trzody chlewnej, jako zagadnienie wymagające odrębnej dyskusji. Nasze uwagi na ten temat miały raczej charakter marginesowy i nie stanowiły zasadniczego tematu opracowania. Dlatego ograniczamy się do problemów metodycznych.

Krytyka sprowadza się do zaprzeczenia poprawności eliminowania tendencji rozwojowej w przypadku poszukiwania współzależności między dwoma zjawiskami dynamicznymi. Abstrahując już od faktu, iż w statystycznej literaturze sprawa ta została od dawna wyjaśniona i nie budzi wątpliwości, należy przypomnieć, że szczególnie gdy ma się do czynienia z korelacją prostoliniową (jak to zachodzi w obu publikacjach) przyjęcie sugestii autorów krytyki daje w efekcie prawie zawsze wysoki współczynnik korelacji (dodatni lub ujemny).

Eliminacja tendencji jest konieczna nawet wtedy, kiedy dwa szeregi czasowe przebiegają równolegle. Równoległość bowiem linii tendencji nie przesądza identycznego stopnia współzmienności, stanowiącego istotę związku korelacyjnego. Jest tylko jeden wyjątek od tej reguły — jeżeli szeregi zmiennych nie wykazują w ogóle tendencji. W przeciwnym razie **jedyną dopuszczalną drogą jest uchwycenie zgodności lub niezgodności zmian przebiegu dwu szeregów czasowych — a o tym właśnie informują odchylenia od trendu.** Jeżeli z pewnych względów (np. o ile wydaje się, że tendencja nie oddaje należycie związku logicznego) wielkości uzyskanych z obserwacji nie chcemy wyrażać w stosunku do trendu, wówczas dysponujemy metodą pierwszych różnic (first differences). Tej metody należało ewentualnie użyć dla sprawdzenia słuszności wysuniętej przez nas hipotezy.

Tymczasem autorzy, stwierdziwszy dość autorytatywnie, iż postępowanie nasze jest nieprzydatne, sugerują z kolei swoje dość oryginalne rozwiązanie. Mianowicie proponują „skorelowanie między sobą współczynników kierunkowych trendów..., tempa ich dynamiki”. **Ten sposób postępowania jest nie do przyjęcia, dlatego że:**

po pierwsze — równanie tendencji rozwojowej prostoliniowej, jak i równanie regresji prostoliniowej są równaniami pierwszego stopnia, w których do współczynnika kierunkowego dochodzi się w sposób identyczny, a zależności pomiędzy poszczególnymi wyrazami równań z punktu widzenia formalno-matematycznego są te same. Wiemy dobrze, że pomiędzy współczynnikiem korelacji a współczynnikiem regresji istnieje ścisła zależność wyrażona wzorami:

$$b_{yx} = r \frac{\sigma_y}{\sigma_x}$$

$$b_{xy} = r \frac{\sigma_x}{\sigma_y}$$

$$r = \sqrt{b_{yx} \cdot b_{xy}}$$

Zależność ta pozostaje w mocy i w przypadku współczynnika kierunkowego równania trendu, mimo że jego treść merytoryczna jest inna. Jeśli o tym nie pamiętamy, to z punktu widzenia formalno-matematycznego popełniamy błąd zasadniczy, który podważa podstawy, na jakich została zbudowana technika rozwiązania. Przecoczenie tego jest tym dziwniejsze, że równocześnie autorzy krytyki, aby obliczyć błąd standardowy oceny (zwany przez nich średnim odchyleniem od trendu) niepotrzebnie sięgają do odchylenia standardowego (str. 15), to znaczy dostrzegają podobieństwa formalne między pewnymi wzorami. W tym jednak przypadku należało od razu posłużyć się wzorem:

$$S = \sqrt{\frac{\sum (X - X')^2}{N}}$$

gdzie: S — błąd standardowy
 X — wartości empiryczne (z obserwacji)
 X' — wartości teoretyczne (obliczane z równania trendu)
 N — liczebność ogólna

po drugie — współczynnik kierunkowy jest bardzo niedoskonałym miernikiem tempa dynamiki (do tego służą przede wszystkim wskaźniki dynamiki), mówi bowiem tylko o przeciętnym wzroście zjawiska (na daną jednostkę czasu) w badanym okresie. W dodatku tym gorzej nas informuje, im rozrzut punktów wokół prostej jest większy. Przy sposobności należy wyjaśnić, że czymś zupełnie różnym jest przyjęcie jednostki czasu (najczęściej rocznej, kwartalnej, lub miesięcznej), a czymś zgoła odmiennym okres badań (który przy tendencji rozwojowej nie powinien być krótszy niż 8—10 lat). Przyjęcie jednostki czasu (takiej czy innej) zależy od typu badań. Nie musi ono wpływać na zniekształcenie zmian, o ile mieszczą się one w okresie badań. Może natomiast zachodzić potrzeba obliczania korelacji z **przesunięciem w czasie**.

po trzecie — w przypadku tendencji krzywoliniowej (równania wyższych stopni) sposób podany przez autorów całkowicie zawodzi (choćbyśmy nawet założyli, że matematycznie jest poprawny); a przecież nie można w każdym przypadku posługiwać się równaniami pierwszego stopnia.

po czwarte — autorzy krytyki zbadali omawianą współzależność w aspekcie przestrzennym, województwami, by dojść do $r = -0,59$. Nie sprawdzając tego współczynnika żadnym testem stwierdzają, iż jest on „istotny”. Przyjmijmy w dalszym ciągu, że jest on poprawnie obliczony i rzeczywiście istotny. Wtedy mówi nam nie o konkurencyjności bydła i trzody względem paszy, ale o efekcie różnic wynikających z ogólnych warunków panujących w danym województwie: tradycji hodowlanych, kierunku produkcyjnego, struktury gospodarstw itp. Odpowiadałby więc na nieco odmiennie pytanie (przy poczynionych błędnych założeniach autorzy oczywiście inną drogą pójść nie mogli, gdyż tylko w ten sposób otrzymali potrzebną dla obliczenia korelacji liczbę współczynników kierunkowych).

Jeżeli autorzy krytyki odczuwali tak wielki brak zaufania do słuszności eliminowania trendu, a nie chcieli sięgnąć do metody pierwszych różnic, to już mniejsze zastrzeżenia budziłoby szukanie korelacji między indeksami dynamiki, choć i tego postępowania bynajmniej nie można uważać za poprawne.

Niejasne jest także sformułowanie: „surowy trend”. Nie ma takiego terminu w statystyce. Mówiąc o współzależności w stosunku do trzeciego wspólnego czynnika autorzy dotykają zagadnienia tzw. korelacji pozornej. O ile taka występuje, wtedy o współzależności nie można mówić. Po prostu trzeba było stwierdzić, że bezpośredniej zależności między pogłowiem bydła i trzody doszukać się nie da. W ten sposób nie byłoby też dyskusji na poruszony temat.

Wypada z zalem stwierdzić, że omawiana publikacja ujawnia z punktu widzenia statystycznego szereg braków. Nie wnikając już w **swoistą terminologię** statystyczną przyjętą przez autorów, ani też w **swoiste charakterystyki** przez nich stosowane (np. średni roczny przyrost jako stosunek współczynnika kierunkowego trendu do średniego plonu w badanym okresie, wyrażony w procentach) trzeba przykładowo zwrócić uwagę na niektóre nieścisłości:

1. Korelacja doskonała (zwana przez autorów całkowitą) nie jest zależnością funkcyjną (s. 8). Mówiąc popularnie, zależność funkcyjna występuje wtedy, gdy jednej wartości zmiennej niezależnej odpowiada jedna i tylko jedna wartość zmiennej zależnej. Zależność korelacyjna występuje zaś wtedy, gdy jednej wartości zmiennej niezależnej odpowiada kilka wartości zmiennej zależnej, które ujmują się jedną charakterystyką, tzn. średnią. O ile ta średnia zmienia się całkowicie zgodnie w stosunku do zmian odpowiedniej zmiennej niezależnej, to mamy wówczas do czynienia z korelacją doskonałą dodatnią.

2. Przy badaniu efektywności nawożenia (s. 26) konieczną rzeczą było zastosowanie korelacji wielorakiej i cząstkowej. Jeżeli zdecydowano się już na dwie zmienne, należało przynajmniej zastosować indeks korelacji lub stosunek współzależnościowy; zachodzi bowiem duże prawdopodobieństwo, że mamy do czynienia z zależnościami krzywoliniowymi.

3. Dla obliczenia tendencji rozwojowej przyjęty został zbyt krótki okres badań (6 lat), stąd wielkości poszczególnych lat zbyt mocno wpływają na wyniki. Zauważyli to zresztą i sami autorzy opracowania.

Na zakończenie trudno powstrzymać się od wyrażenia jeszcze raz żalu, iż tak potrzebna publikacja nie została przed oddaniem jej do druku przeglądnięta przez statystyka.