Henryk Płudowski

Badanie współzmienności cech przy pomocy funkcji potęgowej

Annales Universitatis Mariae Curie-Skłodowska. Sectio H, Oeconomia 13-14, 87-93

1979-1980

ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN — POLONIA
VOL. XIII/XIV, 6 SECTIO H 1979/1980

Międzywydziałowy Instytut Ekonomiki i Organizacji Rolnictwa AR w Lublinie

Henryk PŁUDOWSKI

Badanie współzmienności cech przy pomocy funkcji potęgowej

Исследование ковариантности признаков при помощи степенной функции

Examination of Characteristics Covariation by Means of the Power Function

WPROWADZENIE

Jednym z podstawowych zagadnień występujących przy analizie związków i zależności między zmiennymi jest wybór odpowiedniego modelu funkcji. Przy rozwiązywaniu tego zagadnienia staramy się wybrać taki model, który najlepiej opisuje zależności między interesującymi nas cechami i posiada najkorzystniejsze charakterystyki statystyczne.[1] Chodzi tu głównie o współczynnik determinacji, który określa stopień wyjaśnienia zmienności jednej cechy przez zmienność innych cech wprowadzonych do modelu funkcji, a całe zagadnienie sprowadza się do badania współzmienności.

Wobec tego regresję stosuje się wówczas, gdy występuje między zmiennymi dostatecznie wysoka korelacja, a ściślej determinacja.[2] Należy jednak zwrócić uwagę, że stopień zdeterminowania jednej cechy przez drugą nie może stanowić jedynego kryterium oceny modelu, gdyż nie jest to wystarczające dla właściwego odzwierciedlenia kształtu zależności. Dotyczy to w szczególności funkcji potęgowej, która ze względu na swoje właściwości matematyczno-analityczne jest często stosowana do badania związków i zależności między zmiennymi. Występują okoliczności, że opi-

[1] A. S. Goldberger: *Teoria ekonometrii*, Warszawa 1972; Z. Pawłowski: *Ekonometria*, Warszawa 1969; H. Płudowski: *Badanie efektywności nawożenia metodą funkcji produkcji*, Puławy 1975.

[2] Współczynnik korelacji podniesiony do kwadratu staje się współczynnikiem determinacji.

sywane przy pomocy omawianej funkcji zależności nie odpowiadają rzeczywistości.

Celem artykułu jest przedstawienie niektórych problemów, jakie mogą wystąpić i jakie autor napotkał podczas posługiwania się korelacją i regresją potęgową.[3] Zagadnienia te zostaną przedstawione na przykładzie pochodzącym z Kombinatu PGR Machnów Nowy.[4] Jest to przedsiębiorstwo wieloobiektowe, położone w południowo-wschodniej części woj. zamojskiego, gospodarujące na dobrych glebach rędzinowych; powierzchnia użytków rolnych wynosi 8 tys. ha.

WYNIKI BADAŃ

We wspomnianym przedsiębiorstwie zestawiono i przeliczono na 1 ha użytków rolnych produkcję końcową brutto (rolniczą) i koszty ponoszone na tę produkcję celem określenia związku i zależności między tymi wskaźnikami. Na podstawie wstępnej analizy danych zauważono, że koszty wzrastały szybciej niż produkcja, co wskazywało na to, że między analizowanymi wskaźnikami występowała zależność nieliniowa, a krańcowe przyrosty produkcji na 1 tys. zł przyrostu kosztów były coraz mniejsze. Zastosowano więc funkcję potęgową, której parametry obliczono metodą najmniejszych kwadratów po zlogarytmowaniu zmiennych, czyli obliczano następujący model funkcji:

$$(ln\ y)' = ln\ a + b\ ln\ x,$$

gdzie: ln — logarytm naturalny,[5]
a — stały parametr funkcji,
b — współczynnik regresji,
y — produkcja końcowa brutto,
x — koszty w tys. zł/ha UR.

W wyniku przeprowadzonych obliczeń uzyskano funkcję wyrażającą się równaniem:

$$(ln\ y)' = -0{,}6456 + 1{,}0406\ ln\ x. \qquad \text{(I)}$$

Zdeterminowanie między logarytmami zmiennych y i x jest bardzo wysokie i wynosi 0,9274. Możemy więc powiedzieć, że zmienność $ln\ y$

[3] H. Płudowski: *Model potęgowej funkcji produkcji w zastosowaniu do badania efektywności nawożenia mineralnego*, „Postępy Nauk Rolniczych", 1976, nr 2.

[4] T. Wierzbicki: *Analiza wykorzystania ziemi na tle rozwoju produkcji zwierzęcej w państwowym przedsiębiorstwie rolniczym na przykładzie Kombinatu PGR Machnów Nowy*, Lublin 1978, AR (maszynopis — praca doktorska).

[5] Można również stosować logarytmy dziesiętne, lecz przy małych liczbach lepiej posługiwać się logarytmami naturalnymi.

została wyjaśniona przez zmienność $ln\ x$ w 92,74%, czyli ścisłość związku jest tutaj bardzo wysoka. Można by sądzić, że dokonano trafnego wyboru modelu funkcji, ale w rzeczywistości tak nie jest, bo współczynnik regresji b jest większy od jedności, co wyklucza malejącą efektywność kosztów, jaką zaobserwowano podczas wstępnej analizy danych.

Korzystając z tego, że $ln\ a = -0{,}6456$, można obliczyć współczynnik a i funkcję napisać w postaci potęgowej:

$$Y' = 0{,}524x^{1{,}0406}. \qquad \text{(II)}$$

W zasadzie funkcje (I) i (II) są tożsame w sensie równań matematycznych, ale do (II) nie odnosi się współczynnik determinacji między logarytmami zmiennych, jaki został obliczony dla funkcji (I). Przy aproksymowaniu funkcji $Y' = ax^b$ mamy na myśli funkcję nieliniową, którą przy pomocy logarytmów transformujemy do funkcji liniowej: $(ln\ y)' = ln\ a + b\ ln\ x$. Przyjmując, że $(ln\ y)' = V$, $ln\ a = c$, $ln\ x = z$, otrzymamy:

$$V = c + bz.$$

Przy obliczaniu b i c klasyczną metodą najmniejszych kwadratów minimalizuje się sumę o wyrażeniu $\Sigma(v_i - c - bz_i)^2$, a nie minimalizuje się $\Sigma(y_i - ax_i^b)^2$. Dlatego metoda estymacji odnosi się do regresji między logarytmami zmiennych, a tym samym współczynnik determinacji (r^2_{vz}) dotyczy funkcji (I) i nie równa się indeksowi[6] determinacji potęgowej (i^2_{vz}). Chcąc zatem ocenić funkcję potęgową, należy obliczyć indeks determinacji potęgowej, który można wyrazić wzorem:

$$i_{yx}^2 = 1 - \frac{\Sigma(y_i - ax_i^b)^2}{\Sigma(y_i - \bar{y})^2} = 1 - \frac{\Sigma(y_i - Y_i')^2}{\Sigma(y_i - \bar{y})^2}$$

W naszym przykładzie indeks determinacji potęgowej wynosi 0,8562 i jest mniejszy od współczynnika determinacji między logarytmami zmiennych o 7,12%. Jest to różnica stosunkowo duża i wskazuje, że obliczony współczynnik determinacji dla funkcji (I) nie powinien być stosowany do oceny ścisłości związku wynikającego z funkcji (II).

Biorąc ponadto pod uwagę, że w naszym przypadku funkcja potęgowa nie wykazywała malejącej efektywności wzrastających kosztów, należy dojść do wniosku, iż zastosowany model funkcji nie jest adekwatny do odzwierciedlenia występującej zależności między kosztami a produkcją. Z tego względu zastosowano funkcję paraboliczną i porównano ją z potęgową.

Na podstawie obliczonych współczynników regresji parabolicznej funkcję można wyrazić równaniem:

---

[6] „Współczynnik” dotyczy determinacji liniowej, a „indeks” determinacji nieliniowej — za T. Marszałkowicz: *Metody statystyczne w badaniach ekonomiczno-rolniczych*, Warszawa 1972.

$$Y' = -2,986 + 1,473x - 0,0333x^2, \qquad \text{(III)}$$

gdzie, podobnie jak w funkcji potęgowej, koszty (x) i produkcję (y) wyrażono w tys. zł/ha UR.

Zależność określona funkcją (III) jest zdeterminowana w 97,89%, a więc o 12,27% wyżej niż funkcją potęgową. Również dyspersja resztowa funkcji parabolicznej jest znacznie mniejsza niż w przypadku funkcji potęgowej. Wynosi ona dla funkcji (III) 524 i dla (II) 1313 zł/ha UR. Z tego wynika, że w rozpatrywanym przykładzie funkcja paraboliczna okazała się bardziej uzasadniona z merytorycznego i statystycznego punktu widzenia.

Na podstawie pierwszej pochodnej funkcji (III) można określić, że krańcowe przyrosty funkcji na 1 tys. zł przyrostu kosztów wyrażają się formułą:

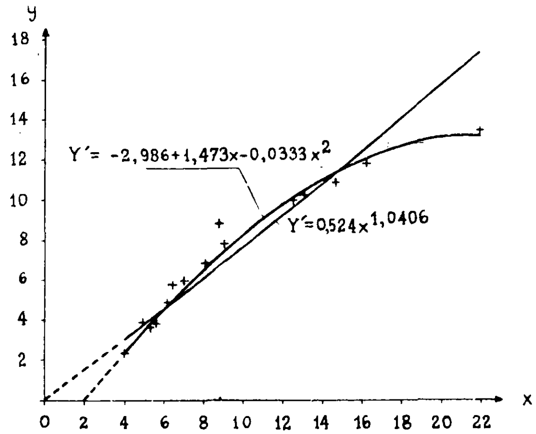
$$\frac{\Delta y}{\Delta x} = 1,473 - 0,0666x.$$

Z porównania omawianych funkcji wynika, że prowadzą one do zupełnie innych wniosków. Na podstawie funkcji potęgowej należałoby stwierdzić, że we wspomnianym przedsiębiorstwie efektywność wzrastających kosztów nie malała, natomiast analiza funkcji parabolicznej wykazuje, że w miarę wzrastania kosztów ich efektywność krańcowa malała, co zgodne jest z rzeczywistością i wcześniejszymi spostrzeżeniami.

Warto też zwrócić uwagę, że współczynnik determinacji między logarytmami zmiennych (0,9274) niewiele różnił się od indeksu determinacji parabolicznej (0,9789), dopiero znaczne różnice wystąpiły przy porównaniu indeksu determinacji potęgowej (0,8562). Wobec tego przy ocenie funkcji pod względem ścisłości związku porównywanie determinacji obliczonej na logarytmach zmiennych z determinacją między naturalnymi wielkościami zmiennych nie jest w pełni uzasadnione i może prowadzić do niewłaściwego wyboru modelu funkcji.

Celem lepszego porównania omawianych funkcji i wyjaśnienia, dlaczego funkcja potęgowa przybrała inny kształt niż oczekiwano, zamieszczono rysunek, na którym współrzędne rozpatrywanych cech zaznaczono krzyżykami.

Z rozkładu współrzędnych wynika, że krzywa paraboliczna lepiej odzwierciedla zależność między kosztami a produkcją w Kombinacie PGR Machnów Nowy niż krzywa potęgowa, której krzywizna jest odwrócona w stosunku do parabolicznej. Krzywa potęgowa bierze swój początek z zerowego punktu układu osi współrzędnych i następnie nawiązuje do rzeczywistego rozkładu obserwacji. W naszym przypadku rozkład współrzędnych nie odpowiadał wspomnianym właściwościom tej funkcji, w wyniku czego powstał zniekształcony obraz w stosunku do rzeczywistości.



Zależność między kosztami (x) a produkcją końcową (y) w tys. złotych na 1 ha UR wyrażona funkcją potęgową i paraboliczną w Kombinacie PGR Machnów Nowy w latach 1960/61—1974/75

The dependence between the costs (x) and the final production (y) in thousands of zlotys per 1 hectare of farmland (UR), expressed with power and parabolic functions in the Collective Farm Combine in Machnów Nowy, in the years 1960/61—1974/75

Krzywa paraboliczna tych właściwości nie posiada i może przecinać osie współrzędnych w dowolnym miejscu, a w obszarze badanej zmienności jest całkowicie przyporządkowana rozkładowi współrzędnych analizowanych cech. Jest więc bardziej elastyczna i lepiej daje się „dopasować" do badanej współzmienności. Jej wadą jest natomiast to, że posiada więcej współczynników strukturalnych i tym samym mniejszą ilość stopni swobody, co może utrudniać statystyczną weryfikację. W naszym przypadku takie trudności nie wystąpiły i wszystkie parametry funkcji potęgowej i parabolicznej są statystycznie istotne przy poziomie prawdopodobieństwa 0,999.

Z porównania omawianych funkcji na przykładzie Kombinatu PGR Machnów Nowy nie wynika jeszcze, że funkcja potęgowa nie nadaje się do badania tego typu zależności i nie o to chodzi. Trzeba natomiast wiedzieć, iż podstawowymi założeniami jej są:

1) asymptotyczność (funkcja permanentnie rosnąca),

2) stała elastyczność względem zmiennych objaśniających,

3) niemożność przybierania wartości ujemnych i wyjście krzywej zawsze z zerowego punktu układu osi współrzędnych,

4) pierwsza pochodna funkcji (przyrosty krańcowe) maleje, lecz nie może przybierać wartości ujemnych.

We wszystkich przypadkach, kiedy te założenia mogą odpowiadać badanej rzeczywistości, funkcja potęgowa może być stosowana i to z dużym powodzeniem. W naszym przykładzie nie odzwierciedlała ona właściwego kształtu zależności ze względu na swój specyficzny przebieg od początku układu osi współrzędnych. Gdyby w podanym przykładzie nie występował ujemny wyraz wolny w funkcji parabolicznej, to analizowaną zależność można byłoby opisać dość dokładnie przy pomocy potęgowego modelu funkcji.

WNIOSKI

Na podstawie przeprowadzonej analizy można sformułować trzy zasadnicze wnioski.

1. Funkcja potęgowa nie powinna być oceniana pod względem ścisłości związku na podstawie współczynnika determinacji pomiędzy logarytmami zmiennych. Oceny takiej można dokonać na podstawie indeksu determinacji potęgowej.

2. Przy wyborze modelu funkcji poza statystyczną weryfikacją współczynników strukturalnych konieczna jest weryfikacja merytoryczna. Z tego względu posługiwanie się metodami statystycznymi wymaga dobrej znajomości badanego zjawiska.

3. Aproksymując model funkcji wskazane jest przeprowadzenie graficznej analizy współzmienności, gdyż to pomaga wyeliminować nieprawidłowości i sformułować poprawne wnioski merytoryczne.

РЕЗЮМЕ

В работе рассматривались вопросы выбора соответствующей модели функции при исследовании ковариантности признаков и оценки точности уравнения, вытекающего из зависимости, которая выражена степенной функцией. Исследования проводились на основе данных, полученных в Государственном сельскохозяйственном комбинате „Махнув-Новы". Анализировалась зависимость, выступающая между издержками производства и конечной продукцией брутто, рассчитанных на 1 га сельскохозяйственных угодий. В результате проведенных исследований установлено, что примененная степенная функция не полностью выражает зависимость, которая выступает между переменными. Оказалось, что в этом случае к разложению координатов изучаемых признаков лучше подходила параболическая модель функции. На основе сравнительного анализа автор пришел к выводу, что степенная функция не должна оцениваться с точки зрения точности уравнения на основе коэффициента детерминации между логарифмами переменных. Такую оценку можно провести при помощи индекса ступенной детерминации. При выборе модели, кроме статистической оценки функции, необходимо также проведение мериторической проверки.

SUMMARY

The article considers the problem of a choice of an appropriate function model when examining the covariation of characteristics and an estimation of the closeness of fit resulting from the dependence expressed by a power function. The analysis was carried out on the basis of the data provided by the Collective Farm Combine in Machnów Nowy. The dependence between the costs and the final gross production per 1 hectare of farmland was analysed. As a result it was shown that the applied power function did not fully reflect the dependence occurring between the variables. It turned out that in the case analysed the parabolic function model fitted more closely the coordinate distribution of the characteristics examined. On the basis of a comparative analysis the author reaches a conclusion that the power function should not be estimated in respect to the closeness of fit of the relation on the basis of the determination coefficient between the logarithms of the variables. Such an estimation may be carried out by means of the power determination index. Besides statistical estimation of the function model, substantial verification is also necessary for making the appropriate choice.