

Ekonometryczny pomiar efektywności ekonomicznej instytucji finansowych. Stochastyczny model graniczny kosztów

Jacek Barbarski*

Nadesłany: 11 sierpnia 2008 r. Zaakceptowany: 10 listopada 2009 r.

Streszczenie

Celem artykułu jest przedstawienie stochastycznego modelu granicznego kosztów jako jednego ze sposobów oceny ekonomicznej efektywności kosztowej instytucji finansowych oraz zastosowanie go do pomiaru efektywności kosztowej oddziałów jednego z polskich banków. Analiza efektywności kosztowej pozwala na zbadanie, czy poszczególne podmioty przy danych cenach czynników produkcji ponoszą minimalny koszt całkowity wytworzenia danego poziomu produkcji.

Do konstrukcji modelu przyjęto dwa zmienne czynniki produkcji (wartość pozyskanych depozytów – kapitał finansowy; wartość wynagrodzeń – kapitał ludzki), jeden czynnik stały (powierzchnia pomieszczeń biurowych – kapitał fizyczny) oraz jeden produkt (wartość udzielonych kredytów).

Rezultatem analizy empirycznej są wskaźniki poziomu efektywności kosztowej oraz rankingi badanej grupy oddziałów w poszczególnych kwartałach. Ze względu na zastosowanie zbyt uproszczonej metody estymacji (skorygowana metoda najmniejszych kwadratów) otrzymane wskaźniki efektywności są zaniżone i za mało dokładne do dalszej analizy. Sporządzone na ich podstawie rankingi są natomiast wiarygodne i mogą być wykorzystane do dalszej oceny i porównań, co należy uznać za istotną korzyść badawczą.

Słowa kluczowe: efektywność ekonomiczna banków, stochastyczny model graniczny kosztów, analiza i ocena banków, efektywność kosztowa, optymalizacja (minimalizacja) kosztów

JEL: C01, C61, D24, G21

* Uniwersytet Ekonomiczny w Krakowie, Katedra Rachunkowości; e-mail: jacek_barbarski@wp.pl.

1. Wstęp

Analiza i ocena działalności podmiotów gospodarczych, w tym instytucji finansowych, stanowią jedno z podstawowych narzędzi wspomagających proces podejmowania w nich decyzji. W praktyce gospodarczej do oceny najczęściej stosuje się metody oparte na analizie wskaźnikowej. Bardzo często okazuje się jednak, że taka ocena jest niewystarczająca. W celu rzetelnego poznania rzeczywistości gospodarczej coraz częściej konieczne staje się poszukiwanie i stosowanie nowych metod, nie tylko w teorii, lecz także w praktyce gospodarczej, głównie z takich dziedzin, jak ekonometria, statystyka czy badania operacyjne. Optymalnym rozwiązaniem byłoby uzupełnianie jednych metod drugimi, a także stosowanie tzw. metod zintegrowanych.

Ocena instytucji finansowych może dotyczyć różnych dziedzin ich działalności. Jednym z najważniejszych obszarów badawczych jest niewątpliwie ocena ekonomicznej efektywności, a jednym ze sposobów pomiaru ekonomicznej efektywności instytucji finansowych jest podejście ekonometryczne oparte na stochastycznych modelach (funkcjach) granicznych. W odróżnieniu od innych sposobów oceny, metody parametryczne mają silne podstawy mikroekonomiczne i statystyczne, a pomiar efektywności opiera się na rozwiązaniu odpowiedniego zagadnienia optymalizacyjnego.

Do parametrycznego pomiaru efektywności początkowo były stosowane modele deterministyczne. Okazały się jednak mało przydatne ze względu na przyjęte założenia ignorujące czynniki o charakterze losowym (brak składnika czysto losowego w specyfikacji tzw. modelu granicznego). W 1977 r., niezależnie od siebie, dwa zespoły badawcze (Aigner, Lovell i Schmidt oraz Meeusen i van den Broeck), na podstawie definicji Farella (1957) zaproponowały do pomiaru efektywności i analizy procesu produkcyjnego stochastyczne modele graniczne, które – jak się okazało – dawały bardziej przekonujące rezultaty.

Dość późno w stosunku do innych dziedzin stosowanie stochastycznych modeli granicznych w bankowości wiązało się z brakiem modelowego ujęcia banku, a także problemem odpowiedniego wyróżnienia do celów badawczych nakładów i efektów instytucji bankowych. Duży wkład w rozwój tych badań wnieśli Sealey i Lindley, którzy w 1977 r. zaproponowali mikroekonomiczny model instytucji finansowej oraz dokonali klasyfikacji czynników produkcji i produktów bankowych (nakładów i efektów). W ten sposób umożliwili zastosowanie (w latach 90. XX w.) stochastycznych modeli granicznych również do pomiaru efektywności instytucji finansowych.

Zgodnie z podejściem zaproponowanym przez Sealeya i Lindleya w badaniach z zakresu bankowości jako czynniki produkcji należy przyjmować kapitał ludzki, kapitał finansowy oraz kapitał rzeczowy. Produktem instytucji finansowej w sensie technicznym jest zestaw usług finansowych świadczonych deponentom i kredytobiorcom. Produktem w sensie ekonomicznym (ostatecznym efektem działalności) są natomiast tylko te usługi (spośród „produktów technicznych”), które mają wyższą wartość w porównaniu z poniesionymi nakładami, a więc produkty finansowe (kredyty i pożyczki) „po stronie aktywów”.

Przeprowadzono już wiele studiów i analiz efektywności technicznej od czasu jej wprowadzenia (w latach 60.), jednak publikowane badania efektywności instytucji finansowych pojawiły się dość późno. Co więcej, do 1993 r. prawie wszystkie opracowania omawiały efektywność amerykańskich banków komercyjnych, a jedynie w kilku zajęto się pomiarem efektywności instytucji

finansowych poza USA¹. Badania efektywności pod względem czasu i zasięgu nie dotrzymały więc kroku zmianom w sektorze usług finansowych (Berger, Hunter, Timme 1993).

Wyniki przeprowadzonych badań nad efektywnością banków, zwłaszcza amerykańskich, nie tylko dostarczają cenne wnioski, lecz także budzą wiele wątpliwości. Dotychczasowe badania wykazały, że nie ma zgodności rankingów efektywności sporządzonych na podstawie różnych metod pomiaru. Dowody dostarczone przez Bergera Hancock i Humphreya (1993) oraz Englisha, Groszkopfa, Hayesa i Yaiswarnga (1993) potwierdzają, że „nieefektywności wyników” są dużo większe niż „nieefektywności nakładów”. Nasuwa się zatem oczywisty wniosek, że konieczne są dalsze badania, szczególnie z zastosowaniem funkcji zysków, a także innych metod uwzględniających nieefektywności wyników.

W krajowej literaturze można znaleźć liczne publikacje z zakresu analizy i oceny działalności polskich banków, jednak badania wykorzystujące metody parametryczne są stosunkowo nowe. Pierwsze badania w Polsce z zastosowaniem stochastycznych modeli granicznych przeprowadzili Osiewalski i Marzec (1998), którzy użyli stochastycznego modelu granicznego do pomiaru efektywności kosztowej oddziałów banku oraz wykorzystali w tym celu wnioskowanie bayesowskie. Badania te zaowocowały licznymi publikacjami z tego zakresu (Osiewalski, Marzec 1996–1997, 1998a, 1998b, 2001; Marzec 1998, 1999; 2002; Osiewalski 2001; Osiewalski, Wróbel-Rotter 1999, 2002), a także praca doktorska obroniona przez Marca (2000). Biorąc pod uwagę liczne zalety oraz cenne wyniki, których dostarczają te metody, istnieje ogromna potrzeba rozwijania tego kierunku badań.

Oprócz stochastycznych modeli granicznych do grupy metod parametrycznych należą m.in.: TFA (*thick frontier approach*) oraz DFA (*distribution-free approach*). W Polsce pod koniec lat 90. zaczęto przeprowadzać badania w celu zmierzenia efektywności kosztowej i dochodowej polskiego systemu bankowego i określenia jej determinant; wykorzystywano do tego metodę DFA. Wyniki powyższych badań zostały zaprezentowane m.in. w pracach: Bednarski i in. (1999; 2001) oraz Nikiel, Opiela (2002).

Przegląd metod analizy i oceny banków zawiera także publikacja książkowa Rogowskiego (1998). Autor wskazał na ogromną potrzebę poszerzenia zakresu zainteresowań badawczych oraz wykorzystywania coraz bardziej zróżnicowanych i wyrafinowanych narzędzi analitycznych do pomiaru i porównań efektywności ekonomicznej banków i oddziałów. Metody pomiaru efektywności, zarówno parametryczne, jak i nieparametryczne, zostały opisane również w publikacjach: Gospodarowicz (2000), Pawłowska (2005).

Celem niniejszego artykułu jest zaprezentowanie głównych założeń, podstaw teoretycznych oraz wybranych sposobów estymacji parametrów granicznej funkcji kosztów dla danych przekrojowych. Za pomocą tego modelu dokonano również ekonometrycznego pomiaru oraz oceny efektywności kosztowej 58 oddziałów jednego z czołowych polskich banków (dla danych z trzech kwartałów).

Artykuł składa się z części teoretycznej oraz empirycznej. W części teoretycznej artykułu kolejno zostały przedstawione następujące zagadnienia: pojęcie efektywności kosztowej oraz jej pomiar, opis krótkookresowej granicznej funkcji kosztów dla danych przekrojowych oraz podstawowe metody estymacji granicznej funkcji kosztów. Część empiryczna zawiera charakterystykę badanej

¹ Poza USA badania efektywności banków i instytucji finansowych były prowadzone głównie w: Kanadzie (1983), Izraelu (1985), Finlandii (1990), Norwegii i Szwecji (1991) oraz w Niemczech (1996).

grupy oddziałów i danych źródłowych oraz wyniki pomiaru oraz oceny efektywności kosztowej za pomocą stochastycznego modelu granicznego kosztów. Do estymacji parametrów powyższego modelu została wykorzystana skorygowana metoda najmniejszych kwadratów.

2. Efektywność (nieefektywność) kosztowa i jej pomiar

W teorii mikroekonomii za pomocą sformalizowanych modeli matematycznych rozważane są trzy podstawowe funkcje celów: funkcja kosztów, funkcja przychodów oraz funkcja zysków. Jednym ze sposobów ich wykorzystania jest analiza efektywności. Podstawowymi narzędziami ekonometrycznej analizy efektywności przedsiębiorstw są stochastyczne modele graniczne. Oprócz pomiaru efektywności umożliwiają one również ocenę, w jakim stopniu firmy realizują określone cele.

W wyniku specyfikacji mikroekonomicznej granicznej funkcji kosztów (*frontier cost function*) faktyczny koszt wytworzenia określonej wiązki produktów zostaje porównany z minimalnym kosztem niezbędnym do wytworzenia takiego samego zakresu usług (produktów). Efektywność kosztowa mierzona jest jako iloraz minimalnego kosztu niezbędnego do wytworzenia danej wielkości produkcji (przy danych cenach czynników) i kosztu rzeczywiście poniesionego przez dany bank (oddział). Analiza efektywności kosztowej pozwala zatem zbadać, czy firma przy danych cenach czynników produkcji ponosi minimalny koszt całkowity wytworzenia określonego poziomu produkcji (Osiewalski, Marzec 1996–1997).

Granica kosztów (*cost frontier*) zostaje określona w wyniku estymacji funkcji kosztów, która przedstawia zaobserwowany koszt w zależności od wielkości produkcji i cen nakładów oraz błędu losowego i zmiennej reprezentującej nieefektywność. Granica ta może być wyrażona w następujący sposób (por. Rogers 1998):

$$C = C(y, w, u_c, v_c) \quad (1)$$

gdzie:

- C – rzeczywiście poniesiony koszt,
- y – wektor wielkości produkcji,
- w – wektor cen nakładów (czynników produkcji),
- u_c – zmienna reprezentująca nieefektywność kosztów,
- v_c – składnik losowy.

Obliczony szacunek efektywności na podstawie tak zdefiniowanej funkcji kosztów uwzględnia zarówno nieefektywność techniczną (zbyt duże zużycie nakładów do wytworzenia określonej wielkości produkcji y), jak i nieefektywność alokacyjną (stosowanie niewłaściwych proporcji każdego rodzaju nakładów i produktów przy danych cenach rynkowych).

Po oszacowaniu danej funkcji granicznej efektywność danego obiektu mierzy się przez porównanie go z obiektem wzorcowym (leżącym na granicy efektywności). Obiektowi o najwyższym poziomie efektywności przypisuje się wartość 1, natomiast pozostałym obiektom przypisuje się wartości z przedziału od 0 do 1. Przykładowo, jeżeli dla danego obiektu wskaźnik efektywności wynosi 0,95, to oznacza, że jest on efektywny w 95% (np. mógłby wytworzyć daną produkcję, wykorzystując 95% poniesionych kosztów), a pozostałe 5% przypisuje się nieefektywności. Na podsta-

wie otrzymanych w ten sposób poziomów efektywności można sporządzić ranking badanej grupy banków (oddziałów), a następnie przeprowadzić analizę porównawczą.

Mikroekonomiczna krótkookresowa funkcja kosztu² przedstawia minimalny koszt niezbędny do uzyskania zadanej wielkości produkcji, przy ustalonych cenach zmiennych czynników produkcji i nakładzie czynnika stałego (zwykle kapitału rzeczowego). Wyraża więc zależność deterministyczną, która w rzeczywistości podlega zakłóceniom. Przedsiębiorstwa ponoszą zazwyczaj wyższy koszt wytworzenia określonego poziomu produkcji niż wynika z mikroekonomicznej funkcji kosztów. Do przyczyn powstawania wyższych kosztów należy tutaj zaliczyć: zakłócenia czysto losowe (w tym również błąd pomiaru kosztów) oraz odchylenia wyłącznie dodatnie, mające źródło wewnątrz przedsiębiorstwa i spowodowane, najogólniej mówiąc, nieoptymalnym zaangażowaniem i wykorzystaniem czynników produkcji. Odchylenia te zostały uznane za miarę nieefektywności kosztowej danego obiektu (Farrell 1957) i informują, w jakim stopniu przedsiębiorstwo ponosi wyższy koszt uzyskania danej produkcji Q , niż wynikałoby to z dostępnej technologii i relacji między cenami czynników. Łącząc deterministyczną, mikroekonomiczną funkcję kosztu (modelowana daną formą funkcyjną) ze złożonym składnikiem losowym, otrzymuje się tzw. stochastyczny model graniczny (*stochastic frontier*) wyprowadzony przez Aignera, Lovella i Schmidta (1977) oraz Meeusena i van den Broecka (1977; por. także Wróbel-Rotter 2001):

$$y_{it} = f(x_{it}, \beta) + \varepsilon_{it}, \wedge \varepsilon_{it} = u_i + v_{it} \text{ dla } i = 1, \dots, N, t = 1, \dots, T \quad (2)$$

gdzie:

$y_{it} = \ln C_{it}$, y_{it} – logarytm kosztu zmiennego i -tego obiektu w okresie t ,

$f(x_{it}, \beta)$ – założona postać mikroekonomicznej funkcji kosztu,

N – liczba obiektów,

x_{it} – wektor wartości zmiennych objaśniających, będących funkcjami cen czynników, wielkości produkcji oraz nakładu czynnika stałego,

β – wektor parametrów formy funkcyjnej charakteryzującej technologię,

v_{it} – symetryczny składnik losowy (zakłada się, że v_{it} mają niezależne identyczne rozkłady normalne $N(0, \sigma_v^2)$,

u_i – nieujemna zmienna losowa reprezentująca nieefektywność; zakłada się niezależność zakłóceń losowych od nieefektywności i niezależność v_{it} oraz u_i od zmiennych objaśniających.

Mając oszacowaną graniczną funkcję kosztów, efektywność kosztową danego obiektu (banku) CE_i można obliczyć jako iloraz:

$$CE_i = \frac{\hat{C}^{\min}}{\hat{C}^i} \quad (3)$$

gdzie:

$\hat{C}^{\min}$ – minimalny koszt niezbędny do wytworzenia danej wielkości produkcji,

$\hat{C}^i$ – koszt rzeczywiście poniesiony.

² Istota pomiaru efektywności kosztowej w przypadku długookresowej funkcji jest taka sama.

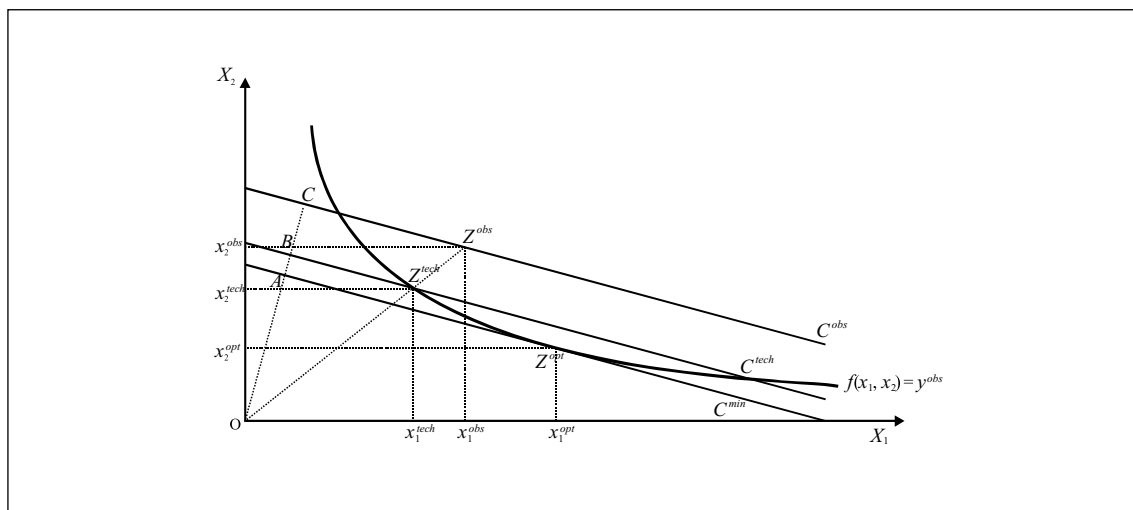
Powyższa relacja pozwala na określenie, w jakim stopniu koszty danego obiektu i (przedsiębiorstwa) są zbliżone do poziomu kosztów, jakie poniósłby obiekt najlepszy z danej próby, gdyby miał wytworzyć taką samą wielkość produkcji (i w takich samych warunkach) jak obiekt analizowany i . Innymi słowy, efektywność kosztowa obiektu i wyraża relację hipotetycznego (oszacowanego na podstawie próby) kosztu, który zostałby poniesiony przez najlepszy bank w próbie w tym samym okresie, gdyby miał wytworzyć takie same wartości produktów i charakteryzować się takimi samymi zmiennymi egzogenicznymi jak obiekt i , do kosztu rzeczywiście poniesionego przez obiekt i . Mając oszacowany wskaźnik efektywności kosztowej (pomijając składnik losowy), poziom nieefektywności danego obiektu oblicza się jako: $1 - CE_i$.

W analizie procesu produkcyjnego można mówić o dwóch podstawowych rodzajach nieefektywności: technicznej i alokacyjnej. Nieefektywność techniczna (*technical inefficiency*) wiąże się z uzyskaniem wielkości produkcji mniejszej od technologicznie możliwej (przy danych nakładach czynników), a nieefektywność alokacyjna (*allocative inefficiency*) wynika z nieprawidłowych proporcji czynników produkcji (krańcowa stopa substytucji nie odpowiada ilorazowi cen rynkowych). Nieefektywność (efektywność) techniczna i alokacyjna stanowią łącznie nieefektywność (efektywność) kosztową (*cost inefficiency*) (Osiewalski, Wróbel-Rotter 1999). Na wykresie 1 została przedstawiona geometryczna dekompozycja efektywności kosztowej. Dla uproszczenia zakłada się, że firma, wytwarzając jeden produkt, wykorzystuje dwa czynniki produkcji przy cenach na ustalonym poziomie.

Na wykresie 1 zbiór wszystkich kombinacji nakładów x_1 oraz x_2 , technicznie niezbędnych do wytworzenia produkcji y^{obs} , przedstawiony jest przez izokwantę $\{(x_1, x_2) : f(x_1, x_2) = y^{obs}\}$, gdzie f – funkcja produkcji}. Gdyby dana firma wykorzystywała optymalne nakłady czynników produkcji $Z^{opt} = (x_1^{opt}, x_2^{opt})$, to produkcję na poziomie y^{obs} uzyskaby po najniższym koszcie C^{min} . W rzeczywistości jej produkcja wynosi y^{obs} przy użyciu nakładów $Z^{obs} = (x_1^{obs}, x_2^{obs})$ i poniesieniu kosztu C^{obs} . Firma jest więc nieefektywna technicznie, ponieważ Z^{obs} leży powyżej izokwenty y^{obs} . Tę samą produkcję y^{obs} można by uzyskać dzięki proporcjonalnej redukcji nakładów (tj. nie zmienia-

Wykres 1

Nieefektywność kosztowa i jej dekompozycja



Źródło: Osiewalski (2001, s. 111).

jąc ich struktury) do punktu Z^{tech} i w ten sposób obniżyć koszty do C^{tech} . Różnica pomiędzy C^{obs} i C^{tech} wyraża koszt nieefektywności technicznej. Nieefektywność alokacyjna polega natomiast na tym, że struktura nakładów Z^{obs} i Z^{tech} nie odpowiada relacji między cenami czynników produkcji, produkcję y^{obs} można bowiem uzyskać taniej (po koszcie $C^{min} < C^{tech}$) przez wykorzystanie optymalnej kombinacji Z^{opt} . Podsumowując, miarą efektywności technicznej jest stosunek odcinków OB/OC, alokacyjnej OA/OB, natomiast kosztowej $OA/OC = (OA/OB) \cdot (OB/OC)$, czyli iloczyn obu mierników cząstkowych (Marzec, Osiewalski 1996–1997).

3. Krótkookresowa graniczna funkcja kosztów – model dla danych przekrojowych

Firma dążąca do maksymalizacji zysku ma do rozwiązania dwa problemy. Pierwszy dotyczy wyboru takiej ilości nakładów czynników produkcji (reprezentowanych przez wektor $x = (x_1, \dots, x_H) \in R^H_+$), która zapewnia minimalny koszt wytworzenia określonej wielkości produkcji Q . Drugi problem wiąże się z wyborem i realizacją takiej wielkości produkcji oraz jej struktury, która pozwala na uzyskanie maksymalnego przychodu ze sprzedaży (przy najniższym koszcie), a w efekcie maksymalizację zysku. Zagadnienie maksymalizacji zysku można więc podzielić na dwa etapy: minimalizację kosztów produkcji (przy danym poziomie produkcji Q) oraz maksymalizację przychodów ze sprzedaży. Należy pamiętać, że problem maksymalizacji zysku polega na znalezieniu takiego poziomu produkcji, w przypadku którego różnica pomiędzy przychodami ze sprzedaży a całkowitymi kosztami wytworzenia jest największa.

Podstawą zbudowania mikroekonomicznego modelu opisującego funkcjonowanie danej firmy jest określenie jej celów działania oraz warunków ograniczających tę działalność. Wśród zazwyczaj wielu celów realizowanych przez każdą firmę podstawowym jest maksymalizacja zysku. Aby ten cel osiągnąć, przedsiębiorstwo określa m.in. rozmiary produkcji, jej strukturę, wybiera czynniki produkcji, napotykać w swojej działalności wiele ograniczeń narzucanych przez otoczenie (ze strony przyrody, dostępnych technologii wytwarzania itp.). Przedsiębiorstwa działają również w różnym otoczeniu rynkowym. Mikroekonomiczny model przedsiębiorstwa opiera się zatem na określonych założeniach realizacji przez firmę celu, którym najczęściej jest maksymalizacja zysku.

Rzeczywistość, nakładając ograniczenia techniczne na przedsiębiorstwo, sprawia, że tylko niektóre kombinacje nakładów są możliwe w celu uzyskania określonej wielkości produktu (produktów), a zatem musi ono ograniczać się wyłącznie do technicznie wykonalnych planów produkcyjnych. Zbiór wszystkich kombinacji nakładów i wyników, który obejmuje technicznie wykonalne sposoby produkcji, nazywany jest zbiorem produkcyjnym, natomiast funkcja opisująca brzeg tego zbioru jest funkcją produkcji (Varian 1999). Zależność pomiędzy ponoszonymi nakładami a wytwarzaną produkcją ma charakter przyczynowo-skutkowy, tzn. przyczyną jest wykorzystanie określonych czynników produkcji, skutkiem zaś uzyskanie odpowiedniej wielkości produkcji. Najbardziej znaną postacią funkcji produkcji jest funkcja typu Cobba-Douglasa.

Problem wyboru przez firmę ilości czynników technicznie niezbędnych do wytworzenia danej wielkości produkcji Q i gwarantujących poniesienie minimalnego kosztu rozwiązywany jest za pomocą formuły minimalizacji kosztów, którą dla ogólnego przypadku można zapisać jako:

$$\min_{x_1, \dots, x_n} (w_1 \cdot x_1 + w_2 \cdot x_2 + \dots + w_n \cdot x_n) \quad \text{przy warunku: } f(x_1, \dots, x_n) = Q \quad (4)$$

gdzie:

$x_1, \dots, x_n$ – czynniki produkcji,

$w_1, \dots, w_n$ – ceny czynników produkcji,

Q – określony (dany) poziom produkcji.

Rozwiązanie problemu minimalizacji kosztów, czyli znalezienie minimalnych kosztów niezbędnych do wytworzenia określonego poziomu produkcji, będzie zależało od cen $w_1, \dots, w_n$ oraz Q , a zatem można go zapisać jako $c(w_1, \dots, w_n, Q)$. Funkcja $c(w_1, \dots, w_n, Q)$ określana jest w mikroekonomii jako funkcja kosztów i wyraża minimalny koszt wytworzenia danego poziomu (wielkości) produkcji Q , kiedy ceny czynników produkcji są dane i wynoszą $(w_1, \dots, w_n)$. W geometrycznej interpretacji rozwiązanie problemu minimalizacji kosztu polega na wyznaczeniu takiego punktu na izokwancie (powierzchni jednakowej produkcji), któremu odpowiada możliwie najniższa izokosta (hiperpłaszczyzna jednakowego kosztu) (Varian 1999).

Rozróżnienie pomiędzy stałymi i zmiennymi wartościami nakładów prowadzi do podziału analizy na krótki i długi okres. W krótkim okresie wielkość wykorzystania niektórych czynników jest stała (w analizie krótkookresowej zakłada się, że przynajmniej jeden z czynników jest stały), co powoduje, że produkcja może wzrastać jedynie w wyniku użycia większej ilości czynników zmiennych. W długim okresie wszystkie czynniki produkcji podlegają zmianie i są traktowane jako zmienne. Krótkookresowa funkcja kosztów wyraża minimalny koszt wytworzenia określonej wielkości produkcji przy co najmniej jednym stałym czynniku produkcji i zmianie pozostałych czynników, natomiast długookresowa funkcja kosztów – w przypadku zmiany wszystkich wykorzystanych czynników produkcji. Należy też dodać, że zdefiniowanie krótkiego i długiego okresu jest sprawą umowną i zależy głównie od specyfiki branży oraz rodzaju prowadzonej analizy.

Dla uproszczenia w analizie procesu produkcji często przyjmuje się, że firma wytwarza jeden produkt Q (bądź produkcję wyrażoną w postaci jednego agregatu mierzzonego w jednostkach pieniężnych), wykorzystując H czynników produkcji reprezentowanych przez wektor $x = (x_1, \dots, x_H) \in R^H_+$. Wówczas taki zbiór $Z = \{x, Q\}$, że z wektora czynników produkcji można otrzymać wielkość produkcji Q , określany jest zbiorem możliwości produkcyjnych (*production possibility set*), a funkcja opisująca brzeg tego zbioru nazywana jest mikroekonomiczną funkcją produkcji lub graniczną funkcją produkcji (*frontier production function*). Funkcja ta wyraża maksymalną produkcję Q możliwą do uzyskania z danych czynników produkcji $x_1, \dots, x_H$ i jest przedmiotem szczególnego zainteresowania firm. Plan produkcji określony na podstawie granicznej funkcji produkcji uznaje się za najbardziej efektywny, ponieważ pozwala uzyskać największą możliwą produkcję z danej ilości czynników. Z własności zbiorów produkcyjnych $\{x, Q\}$ wynika, że funkcja produkcji jest niemalejąca ze względu na czynniki produkcji i wklęsła (Varian 1999).

W przypadku rozważania G produktów firmy ($G \geq 1$) analogonem funkcji produkcji jest funkcja transformacji $T: R^{G+H}_+ \rightarrow R$ taka, że $T(Z^*) = 0$, gdzie Z^* oznacza (technologicznie) efektywny plan produkcyjny, czyli taką kombinację produktów $Q_1, \dots, Q_G$ i czynników produkcji $x_1, \dots, x_H$, że nie jest możliwe uzyskanie większych wielkości produktów przy tych samych nakładach czynników lub wytworzenie tej samej wielkości produktów przy mniejszych nakładach czynników (Marzec 2000).

Jak już wspomniano, zagadnienie minimalizacji kosztów można rozpatrywać dla jednego lub wielu produktów, w krótkim lub długim okresie. Przykładowo, gdy w długim okresie rozważa się jeden produkt Q i H zmiennych czynników produkcji $x_1, \dots, x_H$ o ustalonych cenach $w_1, \dots, w_H$, problem minimalizacji kosztów polega na znalezieniu takich ilości czynników produkcji, dla których wytworzenie danego poziomu produkcji $Q = f(x_1, \dots, x_H)$ odbywa się przy możliwie najniższym koszcie całkowitym $TC = (x_1 \cdot w_1 + \dots + x_H \cdot w_H)$. Minimalny koszt całkowity $c(w_1, \dots, w_H, Q)$ niezbędny do osiągnięcia pożądanego poziomu produkcji Q w takim przypadku jest funkcją cen wszystkich czynników produkcji (zmiennych) $w_1, \dots, w_H$ i wielkości produkcji Q . Nakłady czynników produkcji, które zapewniają firmie minimalny koszt całkowity, tj. nakłady $x_1^*(w_1, \dots, w_H, Q), \dots, x_H^*(w_1, \dots, w_H, Q)$, określane są jako warunkowe funkcje popytu na czynniki i zależą również od cen wszystkich czynników i poziomu produkcji. Otrzymana funkcja kosztu $c(w_1, \dots, w_H, Q)$, odzwierciedlająca minimalny koszt całkowity wytworzenia danego poziomu produkcji Q przy zmianie wszystkich czynników produkcji nazywana jest długookresową funkcją kosztu. Funkcja kosztu jest dualnym (wobec funkcji produkcji) opisem technologii. Mając dane ceny $w_1, \dots, w_H$ oraz technologię opisaną przez funkcję produkcji $Q = f(x_1, \dots, x_H)$ lub funkcję kosztu, można wykorzystać dwa sposoby wyznaczenia optymalnych nakładów czynników produkcji do wyprodukowania ustalonej wielkości Q (Wróbel-Rotter 2001):

- pierwszy opiera się na funkcji produkcji, na podstawie której za pomocą metod optymalizacji można znaleźć funkcje warunkowego popytu na czynniki produkcji,
- drugi wykorzystuje funkcję kosztu, na podstawie której po zróżniczkowaniu względem cen zmiennych czynników produkcji otrzymuje się optymalne nakłady czynników produkcji.

Przedmiotem analizy w niniejszym artykule jest krótkookresowa funkcja kosztu, a więc taka, w której zakłada się, że przynajmniej jeden z czynników produkcji nie podlega optymalizacji (co najmniej jeden z czynników jest stały). Jeżeli przyjmiemy, że np. H -ty czynnik się nie zmienia, wówczas minimalny koszt wytworzenia G produktów (o wielkości $Q_1, \dots, Q_G$) przy zmianie jedynie czynników podlegających optymalizacji (czynników zmiennych: $(x_1, \dots, x_{H-1})$) jest rozwiązaniem zagadnienia (Marzec 2000):

$$\min_{x_1, \dots, x_{H-1}} (x_1 \cdot w_1 + \dots + x_{H-1} \cdot w_{H-1} + \tilde{x}_H \cdot w_H) ; \text{ przy warunku } T(Q_1, \dots, Q_G, x_1, \dots, x_{H-1}, \tilde{x}_H) = 0 \quad (5)$$

czyli:

$$\tilde{x}_H \cdot w_H + \min_{x_1, \dots, x_{H-1}} (x_1 \cdot w_1 + \dots + x_{H-1} \cdot w_{H-1}) ; \text{ przy warunku } T(Q_1, \dots, Q_G, x_1, \dots, x_{H-1}, \tilde{x}_H) = 0 \quad (6)$$

gdzie:

$\tilde{x}_H$ – zaangażowanie czynnika stałego, niepodlegającego optymalizacji.

Przy tak sformułowanym zagadnieniu koszt całkowity w przypadku zaangażowania optymalnych czynników produkcji $x_1^*, \dots, x_{H-1}^*$ zależy od cen wszystkich czynników $w_1, \dots, w_H$, wielkości produktów $Q_1, \dots, Q_G$ oraz nakładów czynnika stałego i nosi nazwę krótkookresowej funkcji kosztu całkowitego (*short run total cost*). Na poziom kosztu całkowitego $c^s(w_1, \dots, w_H, Q_1, \dots, Q_G, \tilde{x}_H)$ składają się: wielkość kosztu zmiennego $VC = c_v^s(w_1, \dots, w_{H-1}, Q_1, \dots, Q_G, \tilde{x}_H)$ równa $(w_i \cdot x_i^*(w_1, \dots, w_{H-1}, Q_1, \dots, Q_G, \tilde{x}_H))$, oraz wielkość kosztu stałego, wynosząca $c_f^s(\tilde{x}_H, w_H) = \tilde{x}_H \cdot w_H$. Krótkookresowa funkcja kosztu zmiennego (*short run variable cost*) jest zatem funkcją cen czynników zmiennych, wielkości produktów oraz nakładów czynników stałych (nie występuje w niej cena czynnika stałego) (Marzec 2000).

Zgodnie z mikroekonomiczną teorią firm krótkookresowa funkcja kosztu zmiennego jest nie-malejąca, jednorodna stopnia jeden i wklęsła ze względu na ceny czynników zmiennych, niemalejąca ze względu na wielkości produktów, nierosnąca i wypukła ze względu na nakłady czynników stałych (Marzec 2002).

Koszt graniczny, gdy wszystkie firmy mają swobodny dostęp do tej samej technologii, reprezentuje minimalny koszt ponoszony przy danych cenach czynników i danym (zrealizowanym) poziomie produkcji. Odchylenie *in plus* od granicznej funkcji kosztów może być interpretowane jako błąd pomiaru, który ze swej istoty może mieć dowolny znak (jest zmienną symetryczną względem zera), lub nieefektywność, będącą zawsze zmienną nieujemną (Marzec, Osiewalski 1996–1997). Mikroekonomiczna krótkookresowa funkcja kosztów jest więc zależnością graniczną wyrażającą minimalny koszt wytworzenia danej produkcji Q przy ustalonych cenach czynników produkcji i nakładach czynnika stałego (Osiewalski, Wróbel-Rotter 2002).

W przypadku danych przekrojowych najprostszy jednorównaniowy model stochastycznej granicznej funkcji kosztów (długookresowej) może być zapisany jako (Marzec, Osiewalski 1996–1997):

$$TC_i = f(Q_i, W_i) \exp(\varepsilon_i) \quad (7)$$

gdzie:

$i = 1, \dots, N$ – liczba firm,

TC_i (Total Cost) – obserwowany koszt całkowity i -tej firmy,

Q_i – wektor $G \times 1$ produktów i -tej firmy,

W_i – wektor $H \times 1$ cen czynników produkcji dla i -tej firmy,

$f(Q_i, W_i)$ – graniczna funkcja kosztów.

Zakłada się, że $\varepsilon_i = u_i + v_i$ oraz u_i i v_i są niezależnymi zmiennymi losowymi, przy czym:

– u_i są to składniki wyrażające nieefektywność badanych firm, o rozkładach niezależnych i wartościach wyłącznie nieujemnych ($u_i \geq 0$, a zatem $E(u_i) > 0$);

– v_i są to symetryczne składniki losowe wyrażające wpływ czynników przypadkowych bądź błędów w pomiarze kosztów; zakłada się, że mają niezależne, identyczne rozkłady o zerowej wartości oczekiwanej, np. $v_i \sim N(0, \sigma_v^2)$.

W literaturze proponuje się różne postacie analityczne funkcji służące do aproksymacji nieznanej funkcji kosztu. W badaniach empirycznych, w tym także z zakresu efektywności kosztowej najczęściej stosuje się tzw. giętkie formy funkcyjne (*flexible functional form*). Ich przykładem jest funkcja translogarytmiczna (*translog cost function*), którą zaproponowali Christensen, Jorgenson i Lau (1971). Otrzymuje się ją – dla logarytmu kosztów ($\ln TC$) – przez aproksymację II rzędu dowolnej funkcji przynajmniej trzykrotnie różniczkowalnej w otoczeniu pewnego (dobranego arbitralnie) punktu. Jej postać w przypadku kosztów całkowitych jest następująca (Marzec, Osiewalski 1996–1997):

$$\begin{aligned} \ln TC_i = & \alpha_0 + \sum_{g=1}^G \beta_g^{(1)} \cdot \ln Q_{i,g} + \sum_{h=1}^H \beta_h^{(2)} \cdot \ln W_{i,h} + \sum_{g=1}^G \sum_{h=1}^H \beta_{g,h}^{(3)} \cdot \ln Q_{i,g} \cdot \ln W_{i,h} + \\ & \sum_{g=1}^G \sum_{j=1}^G \beta_{g,j}^{(4)} \cdot \ln Q_{i,g} \cdot \ln Q_{i,j} + \sum_{h=1}^H \sum_{j=1}^H \beta_{h,j}^{(5)} \cdot \ln W_{i,h} \cdot \ln W_{i,j} + \varepsilon_i \end{aligned} \quad (8)$$

gdzie wobec ε_i czyni się założenia takie jak w równaniu (7).

Jednorodność funkcji kosztów ze względu na ceny czynników produkcji nakłada następujące restrykcje na parametry równania (8) (Marzec, Osiewalski 1996-1997):

$$\begin{aligned}\sum_{h=1}^H \beta_h^{(2)} &= 1 \\ \sum_{h=1}^H \beta_{g,h}^{(3)} &= 0 \quad \text{dla } g = 1, \dots, G \\ \sum_{j=1}^H \beta_{h,j}^{(5)} &= 0 \quad \text{dla } h = 1, \dots, H.\end{aligned}$$

Ponadto zgodnie z twierdzeniem Younga³ $\beta_{g,j}^{(4)} = \beta_{g,j}^{(4)}$ i $\beta_{h,j}^{(5)} = \beta_{j,h}^{(5)}$, co dodatkowo pozwala ograniczyć liczbę swobodnych parametrów.

Jeżeli w równaniu (8) zostaną uwzględnione powyższe warunki symetrii oraz koszt całkowity (TC), a ceny W_h ($h = 1, \dots, H$) zostaną podzielone przez jedną z nich (np. W_H), to jednorodność względem cen będzie automatycznie spełniona. Model przyjmie wówczas następującą równoważną postać (Marzec, Osiewalski 1996–1997):

$$\begin{aligned}\ln\left(\frac{TC_i}{W_{i,H}}\right) &= \alpha_0 + \sum_{g=1}^G \beta_g^{(1)} \cdot \ln Q_{i,g} + \sum_{h=1}^{H-1} \beta_h^{(2)} \cdot \ln\left(\frac{W_{i,h}}{W_{i,H}}\right) + \sum_{g=1}^G \sum_{h=1}^{H-1} \beta_{g,h}^{(3)} \cdot \ln Q_{i,g} \cdot \ln\left(\frac{W_{i,h}}{W_{i,H}}\right) + \\ &\sum_{g=1}^G \sum_{j=g}^G \beta_{g,j}^{(4)} \cdot \ln Q_{i,g} \cdot \ln Q_{i,j} + \sum_{h=1}^{H-1} \sum_{j=h}^{H-1} \beta_{h,j}^{(5)} \cdot \ln\left(\frac{W_{i,h}}{W_{i,H}}\right) \cdot \ln\left(\frac{W_{i,j}}{W_{i,H}}\right) + \varepsilon_i\end{aligned}\quad (9)$$

Liczba estymowanych parametrów w powyższym modelu, z wyjątkiem wyrazu wolnego α_0 , wynosi: $k = (G + H - 1)(G + H + 2)/2$.

W budowie funkcji kosztu służącej do oceny efektywności banków (lub jego oddziałów) zasadniczym problemem jest przyjęcie odpowiednich, z punktu widzenia teorii mikroekonomicznej, produktów oraz czynników produkcji. Stosując podejście „intermediacyjne” do specyfikacji czynników i produktów banku, do budowy modelu przyjęto dwa zmienne czynniki produkcji, tj. wielkość pozyskanych depozytów i innych środków finansowych ogółem wyrażoną w postaci środków pieniężnych (kapitał finansowy) oraz wartość wynagrodzeń brutto (kapitał ludzki), jeden czynnik stały mierzony powierzchnią pomieszczeń biurowych (kapitał fizyczny) oraz jeden produkt wyrażony wartością udzielonych kredytów. W przypadku danych przekrojowych krótkookresowy model graniczny funkcji kosztu zmiennego przyjmuje następującą postać⁴:

³ Wynika z niego w szczególności, że jeżeli funkcja f jest dwukrotnie różniczkowalna w pewnym punkcie i pochodne drugiego rzędu są ciągłe, to macierz pochodnych drugiego rzędu w tym punkcie jest macierzą symetryczną.

⁴ W badaniach empirycznych wykorzystano tę postać modelu.

$$\begin{aligned} \ln VC_i = & \beta_0 + \beta_1 \ln w_{i,D} + \beta_2 \ln w_{i,L} + \beta_3 \ln Q_i + \beta_4 \ln K_i + \beta_5 \ln w_{i,D} \ln w_{i,L} + \\ & \beta_6 \ln w_{i,D} \ln Q_i + \beta_7 \ln w_{i,D} \ln K_i + \beta_8 \ln w_{i,L} \ln Q_i + \beta_9 \ln w_{i,L} \ln K_i + \\ & \beta_{10} \ln Q_i \ln K_i + \beta_{11} (\ln w_{i,D})^2 + \beta_{12} (\ln w_{i,L})^2 + \beta_{13} (\ln Q_i)^2 + \beta_{14} (\ln K_i)^2 + v_i + z_i \end{aligned} \quad (10)$$

gdzie:

- $\ln VC_i$ – obserwowany poziom kosztu zmiennego (koszty kapitału ludzkiego w postaci wynagrodzeń brutto wraz z innymi narzutami (ZUS, PFRON, PFŚP), koszty kapitału finansowego w postaci odsetek od depozytów i innych pozyskanych środków oraz koszty zaangażowania kapitału fizycznego (rzeczonego) z wyjątkiem nieruchomości i -tego oddziału banku,
- $w_{i,D}$ – cena czynnika finansowego (depozytów i innych pozyskanych środków),
- $w_{i,L}$ – cena pracy (wysokość średniego wynagrodzenia brutto jednego zatrudnionego pracownika),
- Q_i – zaobserwowana wielkość produkcji wyrażona wartością udzielonych kredytów,
- K_i – zaangażowanie czynnika stałego (kapitału fizycznego) mierzonego powierzchnią (w metrach kwadratowych) pomieszczeń biurowych własnych i najmowanych,
- v_i – składnik losowy o rozkładzie symetrycznym,
- z_i – zmienna reprezentująca nieefektywność kosztową.

Na podstawie powyższego stochastycznego modelu granicznego krótkookresowej, translogarytmicznej funkcji kosztu zmiennego został przeprowadzony ekonometryczny pomiar efektywności kosztowej oddziałów jednego z czołowych banków polskich.

4. Ekonometryczna estymacja granicznej funkcji kosztów

Główny problem w ekonometrycznej analizie efektywności sprowadza się do oszacowania parametrów określonej funkcji granicznej (w tym przypadku funkcji kosztów) oraz wskaźników mierzących poziom efektywności kosztowej (bądź jej nieefektywności). Do modelowania granicznej funkcji kosztów można zastosować klasyczne metody ekonometryczne, jak również metody programowania liniowego i kwadratowego. Te ostatnie mają kilka istotnych wad. Do najważniejszych z nich należy zaliczyć dużą wrażliwość na obserwacje nietypowe, a także nieuwzględnianie informacji dotyczących rozkładu zmiennej reprezentującej nieefektywność.

Najprostsze metody estymacji jednorównaniowych stochastycznych modeli granicznych oparte są na metodzie najmniejszych kwadratów (MNK). W przypadku danych przekrojowych, gdy graniczna funkcja kosztów⁵ jest liniowa względem $K + 1$ parametrów, przybiera ona następującą ogólną postać (Marzec 2000):

$$y_i = \alpha_0 + x_i \beta_{(-0)} + v_i + u_i \quad \text{dla } i = 1, \dots, N \quad (11)$$

gdzie:

$\beta_{(-0)} = [\beta_1, \dots, \beta_K]^T$ – wektor K nieznanymi parametrów charakteryzujących technologię (z wyłączeniem wyrazu wolnego α_0),

⁵ Ponieważ $u_i \geq 0$, y_i jest kosztem (lub jego rosnącą funkcją, np. logarytmem) w przypadku funkcji kosztu.

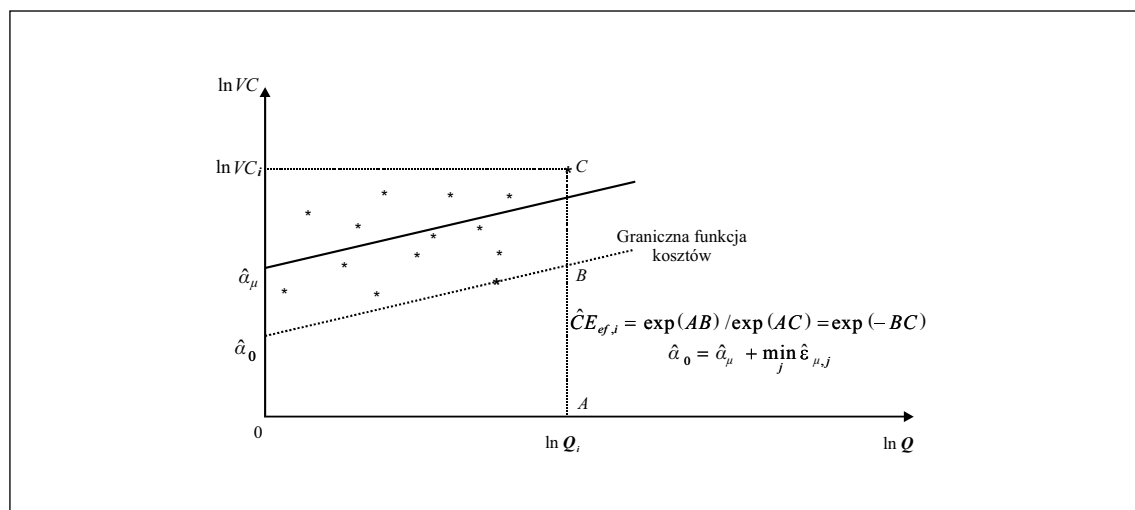
u_i – składniki modelujące potencjalną nieefektywność (v_i oraz u_i są zmiennymi losowymi niezależnymi od siebie).

Ze względu na to, że w modelu (11) występuje N wielkości nieobserwowalnych u_i modelujących potencjalną nieefektywność, a całe wnioskowanie o nich może być oparte na tej samej liczbie obserwacji N , wprowadza się oszczędną parametryzację modelu, zakładając, że składniki u_i mają niezależne od siebie rozkłady ($u_i \perp u_j$) indeksowane przez wspólne parametry. W tym celu należy np. przyjąć, że niezależne nieujemne zmienne losowe u_i mają identyczną wartość oczekiwaną $E(u_i) = \mu$. Trzeba też pamiętać, że suma zmiennych losowych $\varepsilon_i = v_i + u_i$ ma rozkład, który nie jest rozkładem normalnym ani rozkładem symetrycznym. Co najważniejsze, wartość oczekiwana złożonego składnika losowego jest niezerowa: $E(\varepsilon_i) = E(v_i) + E(u_i) = E(u_i) > 0$. Jeżeli wartości oczekiwane zmiennych u_i są identyczne, to przyjmując oznaczenia: $E(u_i) = \mu$, $\alpha_{\mu} = \alpha_0 + \mu$ oraz $\alpha_{\mu,i} = \varepsilon_i - \mu$, otrzymujemy równoważny modelowi (11) zapis, w którym wyrazem wolnym jest α_{μ} , składnikiem losowym zaś $\varepsilon_{\mu,i}$ (Marzec, Osiewalski 1996–1997):

$$y_i = \alpha_0 + \mu + x_i \beta_{(-0)} + v_i + u_i - \mu = \alpha_\mu + x_i \beta_{(-0)} + \varepsilon_{\mu,i} \quad (12)$$

Ponieważ $E(\varepsilon_{\mu,i}) = E(\varepsilon_i) - \mu = E(u_i) - \mu = 0$, więc – przy standardowych założeniach o zmiennych objaśniających – estymator MNK wektora zawierającego α_μ i wszystkie współczynniki $\beta^{(k)}$ funkcji translogarytmicznej ($k = 1, \dots, K$) jest estymatorem (przynajmniej) zgodnym. Uzasadnione jest więc wykorzystanie MNK do estymacji modelu (11), przy czym niezbędna jest korekta oceny wyrazu wolnego, gdyż to estymator $\alpha_\mu = \alpha_0 + \mu$ jest zgodny, a nie estymator α_0 (Marzec, Osiewalski 1996–1997). Zgodność estymatora MNK wszystkich współczynników, poza wyrazem wolnym, może być zatem podstawą różnych wariantów estymacji parametrów i obliczania wskaźników efektywności funkcji granicznych.

Wyznaczanie wskaźnika efektywności kosztów na podstawie danych przekrojowych (skorygowana MNK)



Źródło: Marzec, Osiewalski (1996–1997, s. 72).

Najprostszy wariant znalezienia funkcji granicznej nosi nazwę skorygowanej MNK (*corrected ordinary least squares: COLS*) i został zaproponowany przez Winstena (1957), choć jego odkrycie przypisuje się Gabrielsenowi (1975). Sprowadza się on do estymacji MNK równania (12) oraz wyznaczenia reszt $\hat{\varepsilon}_{\mu,i} = y_i - \hat{y}_i$. Przyjmuje się również, że jedna z badanych firm jest w pełni efektywna, a zatem nieefektywność pozostałych mierzy się w stosunku do niej. W przypadku funkcji kosztów firmę, dla której reszta $\hat{\varepsilon}_{\mu,i}$ jest minimalna, traktuje się jako w pełni efektywną (jako wzorzec efektywności). Minimalna wartość ujemnej reszty $\hat{\varepsilon}_{\mu,i}$ jest oszacowaniem błędu systematycznego $-\mu$. Wykres 2 przedstawia graficzną ilustrację wyznaczania granicznej funkcji kosztów (przyjęto założenia: jeden produkt Q , ceny czynników produkcji oraz nakłady czynników stałych są ustalone dla wszystkich firm).

Graniczna funkcja kosztów (linia przerywana) o wyrazie wolnym $\hat{\alpha}_0 = \hat{\alpha}_\mu + \min_j(\hat{\varepsilon}_{\mu,j})$ jest wyznaczana przez przesunięcie wzdłuż osi rzędnych o wektor $(0; \min_j(\hat{\varepsilon}_{\mu,j}))$ tzw. przeciętnej krzywej kosztów (linia ciągła) oszacowanej za pomocą MNK na podstawie punktów empirycznych. Wykorzystując tak wyznaczoną graniczną funkcję kosztów, można określić miarę względnej efektywności kosztowej ($\hat{CE}_{ef,i}$) jako iloraz minimalnego kosztu niezbędnego (wynikającego z granicznej funkcji kosztów) do kosztu rzeczywiście poniesionego przy tym samym poziomie produkcji. Ze względu na to, że rozważamy logarytm funkcji kosztów, efektywność i -tej firmy na skali $(0, 1)$ mierzy się w następujący sposób (Marzec, Osiewalski 1996–1997):

$$\hat{CE}_{ef,i} = \frac{\exp(\hat{y}_i + \min_j(\hat{\varepsilon}_{\mu,j}))}{\exp(y_i)} = \frac{\exp(\hat{y}_i + \min_j(\hat{\varepsilon}_{\mu,j}))}{\exp(\hat{y}_i + \hat{\varepsilon}_{\mu,i})} = \exp(\min_j(\hat{\varepsilon}_{\mu,j}) - \hat{\varepsilon}_{\mu,i}) \quad (13)$$

Główną zaletą skorygowanej MNK jest prostota obliczeń – do szacowania parametrów stosowana jest wyłącznie MNK, natomiast oceny efektywności w przypadku każdej obserwacji uzyskuje się wprost z reszt MNK. Nie trzeba też zakładać żadnego rozkładu dla zmiennej u_i . Zasadniczą wadą tej metody jest pominięcie składnika losowego ($v_i = 0$), a więc wszystkie odchylenia traktowane są jako nieefektywność.

Drugim sposobem estymacji funkcji granicznych może być korekta wyrazu wolnego o ocenę wartości oczekiwanej złożonego składnika losowego $\varepsilon_i = v_i + u_i$, po uprzednim zastosowaniu MNK do równania (12). W wariancie tym estymacja sprowadza się do oszacowania parametru μ , o który koryguje się $\hat{\alpha}_\mu$ w celu znalezienia $\hat{\alpha}_0$. W metodzie tej nie ignoruje się symetrycznego składnika losowego v_i , ale też nie ma pewności, że po przesunięciu przeciętnej funkcji granicznej wszystkie odchylenia od niej będą miały ten sam znak. Konieczne staje się wówczas zastosowanie innej metody pomiaru indywidualnej nieefektywności.

W celu oszacowania wartości oczekiwanej $E(u_i)$ oraz pozostałych parametrów rozkładu złożonego składnika losowego $\hat{\varepsilon}_{\mu,i}$ stosuje się np. metodę momentów, czyli porównuje się odpowiednią liczbę niezerowych momentów teoretycznych składników losowych $\hat{\varepsilon}_{\mu,i}$ z momentami empirycznymi reszt MNK z równania (12), a następnie dokonuje się korekty wyrazu wolnego tego równania (Marzec 2000):

$$\hat{\alpha}_0 = \hat{\alpha}_\mu - \hat{E}(u_i) = \hat{\alpha}_\mu - \hat{\mu} \quad (14)$$

Omówiona technika estymacji nosi nazwę zmodyfikowanej metody najmniejszych kwadratów i została zaproponowana przez Richmonda (1974).

Do estymacji stochastycznych modeli granicznych Koop i Mullahy (1989) zaproponowali również wykorzystanie uogólnionej metody momentów (*Generalized Method of Moments*). Wymaga ona jedynie, aby zmienna odzwierciedlająca nieefektywność (u_i) posiadała dowolny rozkład ucięty w zerze i określony na dodatniej półosi ($u_i > 0$), natomiast zakłócenie losowe (v_i) miało rozkład symetryczny. Ponadto pierwszych sześć momentów zmiennych u_i i v_i powinno być skończonych, co jednak powoduje, że przy tak ogólnych założeniach nie daje ona zadowalających ocen parametrów.

Oprócz skorygowanej MNK, zmodyfikowanej MNK i uogólnionej metody momentów do estymacji stochastycznych modeli granicznych wykorzystywana jest także metoda największej wiarygodności (MNW). W MNW poszukuje się wartości wektora parametrów α i parametrów struktury stochastycznej, które wektorowi zaobserwowanych w próbie wartości zmiennej objaśnianej $y = (y_1, \dots, y_N)$ gwarantują największą gęstość prawdopodobieństwa. Aby zbudować funkcję wiarygodności, należy poczynić założenia dotyczące rozkładu zmiennej reprezentującej nieefektywność. Podobnie jak w przypadku MOLS najczęściej przyjmowane są rozkłady: ucięty normalny lub wykładniczy. Należy też pamiętać, że utrzymane są w mocy założenia z modelu (11). Po uwzględnieniu faktu, że każdy składnik losowy składa się z dwóch elementów, maksymalizowana jest łączna gęstość wektora $y = (y_1, \dots, y_N)$, którą można otrzymać z łącznej gęstości u_i oraz v_i ($i = 1, \dots, N$). Ze względu na niezależność technicznej nieefektywności od zakłóceń czysto losowych łączną gęstość u_i oraz v_i można przedstawić jako iloczyn gęstości obu zmiennych (odpowiednio $f(v_i)$ i $h(u_i)$). Funkcja gęstości złożonego składnika losowego $v_i = \varepsilon_i - u_i$ dla poszczególnych obiektów wyraża się następującym wzorem (Osiewalski, Wróbel-Rotter 1999):

$$f_*(\varepsilon_i) = \int_0^{\infty} f(\varepsilon_i - u_i) h(u_i) du_i \quad (15)$$

a funkcję wiarygodności można zapisać jako:

$$L(\psi|Y) = \prod f_*(\varepsilon_i) = \prod_{i=1}^N \int_0^{\infty} f(\varepsilon_i - u_i) h(u_i) du_i \quad (16)$$

gdzie:

ψ – wektor parametrów strukturalnych rozpatrywanego modelu, składający się z wektora α i parametrów charakteryzujących rozkłady prawdopodobieństwa u_i i v_i ,

$\varepsilon_i = y_i - (\alpha_0 - x_i \beta_{(-0)})$,

$f(\varepsilon_i - u_i)$ – funkcja gęstości zmiennej $v_i = \varepsilon_i - u_i$,

$h(u_i)$ – funkcja gęstości zmiennej u_i .

Ze względu na swe wady przedstawione podstawowe techniki estymacji stochastycznych modeli granicznych nie pozwalają na precyzyjny pomiar efektywności. Pełne wnioskowanie o charakterystykach procesu produkcyjnego oraz o indywidualnej efektywności firm możliwe jest dopiero na gruncie bayesowskim. Łączy ono w procesie estymacji wstępną wiedzę o analizowanym zjawisku (np. wynikającą z teorii ekonomicznej) oraz informacje zawarte w próbie statystycznej, co przedstawiają van den Broeck i in. (1994).

W badaniach empirycznych ze względu na prostotę obliczeń do estymacji stochastycznego modelu granicznego kosztów zostanie jednak wykorzystana skorygowana MNK.

5. Charakterystyka badanej grupy oddziałów oraz danych źródłowych

Empiryczna analiza efektywności zostanie przeprowadzona na przykładzie oddziałów jednego z polskich banków komercyjnych – nie tylko z powodu braku odpowiednich danych dotyczących poszczególnych polskich banków, lecz także ze względu na znacznie większą porównywalność danych dotyczących oddziałów jednego banku (ta sama technologia stosowana w obrębie jednego banku). Wykorzystane w badaniach dane kwartalne (II, III i IV kwartał)⁶ pochodzą z wybranego jednego roku działalności 58 oddziałów jednego z dużych polskich banków ($T = 3$, $N = 58$). Z próby zostały wyeliminowane oddziały nowo powstałe, a filie połączono z ich oddziałami macierzystymi. Zakres oraz przekrój danych były w dużej części uzależnione od struktury zakładowego planu kont, według którego wygenerowano potrzebne informacje. Sposób ewidencji danych ograniczył liczbę i rodzaj czynników finansowych oraz produktów, które zostały wykorzystane w analizie empirycznej. Z konieczności więc, zarówno depozyty (złotowe i walutowe, od instytucji finansowych i klientów), jak i kredyty (złotowe i walutowe, dla instytucji finansowych i klientów) zostały zagregowane do jednego rodzaju.

Oddziały analizowanego banku mają siedziby w miastach o różnej wielkości, w regionach o różnym charakterze gospodarczym (przemysłowym, turystycznym, rolniczym). Zakres świadczonych usług zależy więc od charakteru lokalnego rynku i wynikających z niego potrzeb klientów. Rozmiary działalności poszczególnych oddziałów można scharakteryzować za pomocą wielkości zgromadzonych środków finansowych, jak również udzielonych kredytów. Z powodu braku rzeczywistych danych przyjęto założenie upraszczające, że oddziały, które mają nadwyżkę pozyskanych środków finansowych nad udzielonymi kredytami (oddziały depozytowe), w całości przeznaczają ją – za pośrednictwem centrali banku – dla innych oddziałów, natomiast oddziały mające niedobór środków finansowych (oddziały kredytowe) pozyskują je – również z centrali – do wysokości wymaganego zapotrzebowania. Założenie takie powoduje, że wielkość zgromadzonych środków finansowych równa się wielkości udzielonych kredytów⁷.

Podstawowy podział wyróżnia oddziały depozytowe, czyli mające nadwyżkę przyjętych lokat (depozytów) nad udzielonymi kredytami, oraz kredytowe, w których z kolei występuje przewaga kredytów. Charakter działalności poszczególnych oddziałów powoduje, że konieczne jest dokonywanie wzajemnych rozliczeń finansowych pomiędzy oddziałami bądź za pośrednictwem centrali banku⁸. Podział oddziałów na depozytowe i kredytowe sprawia, że wynik z działalności operacyjnej oddziałów kredytowych korygowany jest in minus o dodatkowe koszty odsetkowe wynikające z pozyskania brakujących środków finansowych, czyli różnicę pomiędzy wartością kredytów udzielonych przez oddział i zgromadzonych przez niego depozytów, przemnożoną przez cenę pieniądza ustaloną przez centralę banku. Uzyskuje się w ten sposób tzw. skorygowany wynik dzia-

⁶ Pierwszy kwartał nie został uwzględniony z powodu braku odpowiednich danych.

⁷ W przeprowadzonej analizie empirycznej założenia takiego nie uznaje się za błąd merytoryczny.

⁸ W oddziałach depozytowych występuje przewaga kosztów odsetkowych nad przychodami odsetkowymi (są to oddziały typowo kosztowe), natomiast w oddziałach kredytowych występuje sytuacja odwrotna. W celu uzyskania bardziej obiektywnej oceny wyniki ich działalności są korygowane o przepływy międzyoddziałowe.

łałości operacyjnej. Analogicznie, w przypadku oddziałów depozytowych dokonuje się korekty wyniku działalności operacyjnej in plus o dodatkowe przychody odsetkowe uzyskane w wyniku przekazania innym oddziałom nadwyżki środków finansowych, tj. różnicę pomiędzy wartością depozytów posiadanych przez oddział i udzielonych kredytów, przemnożoną przez wewnętrzną cenę „rozliczeniową” pieniądza. W analizie empirycznej dla uproszczenia przyjęto, że w oddziałach kredytowych dodatkowe środki pozyskiwane od innych oddziałów traktowane są łącznie z pozostałymi depozytami i stanowią jeden finansowy czynnik produkcji. Z kolei w oddziałach depozytowych środki przekazywane innym oddziałom zostały połączone z innymi kredytami, w wyniku czego uzyskano jeden produkt.

Stosując w badaniach empirycznych podejście intermediacyjne, jako czynniki produkcji przyjęto:

- a) liczbę pracowników oddziałów – czynnik ludzki (zmienną oznaczono jako L),
- b) zgromadzone depozyty (złotowe i walutowe) oraz pozyskane od innych oddziałów lub centrali środki pieniężne, stanowiące czynnik finansowy (D),
- c) nieruchomości jako podstawowy kapitał fizyczny (K) – czynnik stały,
- d) inne środki trwałe (m.in. sprzęt komputerowy) oraz wartości niematerialne i prawne, stanowiące zmienny kapitał niefinansowy (M).

Przyjęcie dwóch kategorii kapitału fizycznego, K i M , wynika z założenia, że nieruchomości muszą być traktowane w analizie krótkookresowej jako czynnik stały (nie podlegający optymalizacji). Czynnik M można natomiast traktować jako czynnik zmienny, mający swoją cenę. W analizie empirycznej można przyjąć, że cena czynnika M jest jednakowa dla każdego oddziału, ze względu na fakt, że centrala banku w drodze przetargu wybiera dostawców oprogramowania oraz sprzętu (komputerów, drukarek, akcesoriów komputerowych itp.). Podstawowy warunek ekonomicznej regularności, tj. homogeniczność względem cen czynników zmiennych, będzie spełniony, gdy każdą z cen oraz wielkość objaśnianą (koszt, przychód, zysk) wyodrębni się jak we wzorze (9). W rozważanych stochastycznych modelach granicznych przyjęto, że powyższe kategorie ekonomiczne podzielono przez cenę czynnika fizycznego M .

Zaangażowanie poszczególnych czynników produkcji proponuje się mierzyć odpowiednio poprzez:

- a) liczbę zatrudnionych w przeliczeniu na pełne etaty,
- b) wartość zgromadzonych depozytów i innych pozyskanych środków pieniężnych,
- c) powierzchnię pomieszczeń biurowych własnych i wynajmowanych przez oddział,
- d) wartość brutto przyjętych środków trwałych oraz wartości niematerialnych i prawnych.

Produkcja w badaniach empirycznych zostanie natomiast wyrażona jako łączna wartość różnego rodzaju udzielonych kredytów oraz – w przypadku oddziałów depozytowych – środków przekazanych innym oddziałom, która stanowi jeden zagregowany produkt. Wartość pozyskanych depozytów i lokat, jak też udzielonych kredytów w poszczególnych kwartałach wyrażono jako ich średnią wartość według stanu na początek i koniec danego kwartału.

W konstrukcjach stochastycznych modeli granicznych występują ceny określonych powyżej zmiennych czynników produkcji oraz ceny przyjętego zagregowanego produktu. Jako cenę pracy ludzkiej (w_L) przyjęto średnie wynagrodzenie pracownika oddziału wraz z narzutami (ZUS, PFRON, PFŚP, zaliczki na podatek od osób fizycznych) w danym okresie (kwartale). Cenę czynnika finansowego (w_D) stanowi średnia ważona rzeczywistej ceny przyjmowanych lokat i depozytów bankowych (bieżąca cena obowiązująca na rynku lokalnym) oraz wymienionej wcześniej

wewnętrznej ceny rozliczeniowej w danym okresie (kwartale). Wagami wykorzystywanymi do konstrukcji w_D są udziały zgromadzonych depozytów i lokat oraz środków pieniężnych pozyskanych od innych oddziałów (bezpośrednio lub za pośrednictwem centrali). W przypadku oddziałów depozytowych udział środków pozyskanych od innych oddziałów wynosi 0, a więc cena w_D pokrywa się z rzeczywistą ceną depozytów i lokat występującą na rynku. Analogicznie została określona cena produktu w_P .

6. Ekonometryczny pomiar efektywności kosztowej za pomocą granicznej funkcji kosztów

W ekonometrycznej analizie efektywności ekonomicznej oddziałów został wykorzystany stochastyczny model graniczny kosztów dla danych przekrojowych. Do estymacji parametrów powyższego modelu, ze względu na łatwość obliczeń, została natomiast wykorzystana skorygowana metoda najmniejszych kwadratów.

Wyniki obliczeń zaprezentowano w tabeli 1. Po przeprowadzeniu obliczeń dla każdego z okresów (kwartałów) uzyskano współczynniki determinacji na poziomie 0,996. Pomimo wysokich wartości współczynników determinacji oszacowane indywidualne poziomy nieefektywności należy uznać za mało precyzyjne ze względu na uproszczoną metodę estymacji i pomiaru samej efektywności. Otrzymane rankingi oddziałów na podstawie obliczonych wskaźników należy natomiast uznać za wiarygodne.

Uzyskane wyniki wskazują, że w II kwartale średnio 9% kosztu zmiennego poniesionego przez wszystkie oddziały było kosztem nadwyżkowym, nieuzasadnionym ani wielkością produkcji, ani cenami czynników produkcji czy wielkością czynnika stałego (w III i IV kwartale odpowiednio: 11% i 23%).

W zależności od przyjętego kryterium oceny ekonomicznej efektywności (granicznej funkcji kosztów) poszczególne oddziały zajmują różne miejsca w rankingach. Najwyższą efektywnością kosztową w poszczególnych kwartałach charakteryzują się następujące oddziały:

- a) w II kwartale: oddziały o nr.: 41, 36, 33, 47, 48;
- b) w III kwartale: oddziały o nr.: 35, 47, 25, 48, 36;
- c) w IV kwartale: oddziały o nr.: 6, 47, 41, 35, 48.

Z kolei do najmniej „efektywnych kosztowo” oddziałów należą:

- a) w II kwartale: oddziały o nr.: 32, 52, 14, 27, 31;
- b) w III kwartale: oddziały o nr.: 32, 52, 29, 13, 19;
- c) w IV kwartale: oddziały o nr.: 14, 29, 19, 16, 17.

Wszystkie pozostałe wyniki obliczeń (wskaźniki efektywności kosztów oraz rankingi oddziałów) zamieszczono w tabeli 1. Dla uzupełnienia analizy należy dodać, że średni poziom efektywności oddziałów w poszczególnych kwartałach wynosi odpowiednio: 0,910; 0,895; 0,771. Oznacza to, że każdy następny kwartał charakteryzował się spadkiem średniego poziomu efektywności w badanej próbie. W II kwartale średnio 9% kosztów zmiennych poniesionych przez wszystkie oddziały było kosztem nadwyżkowym, nieuzasadnionym wielkością produkcji, cenami czynników produkcji czy wielkością czynnika stałego (w III i IV kwartale odpowiednio: 10,5% oraz 22,9%). Wyraźnie też widać spadek „indywidualnego” poziomu efektywności w zdecydowanej większości

Tabela 1

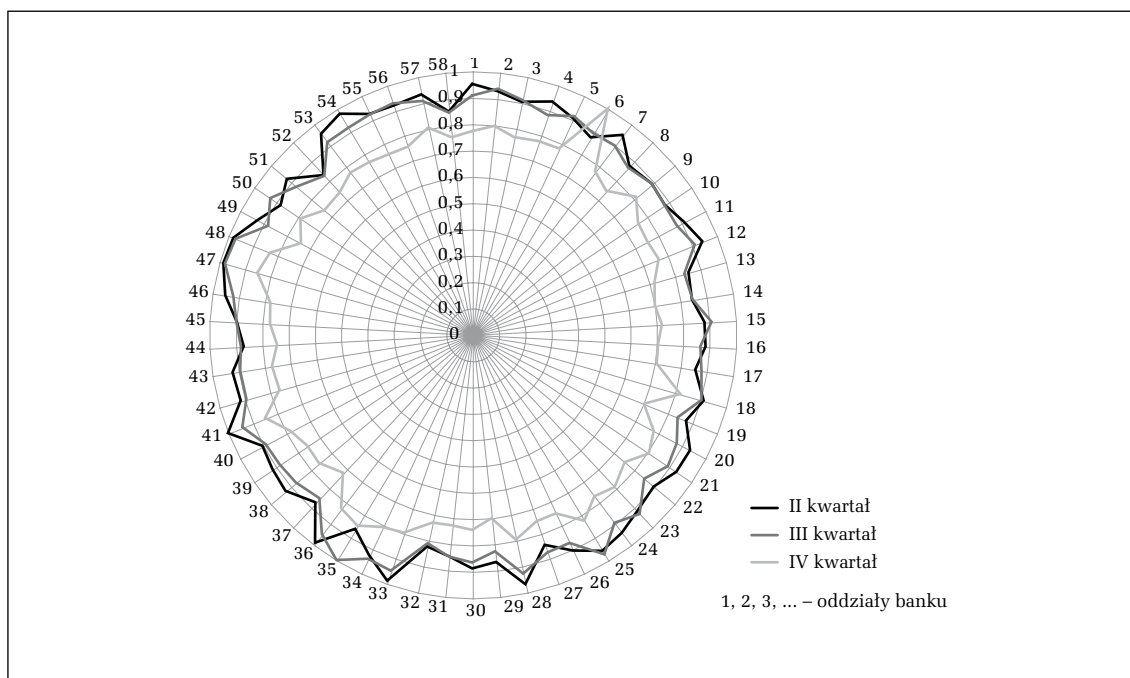
Porównanie ocen efektywności kosztowej oddziałów banku

Nr oddziału	II kwartał	Ranking	III kwartał	Ranking	IV kwartał	Ranking
1	0,951	10	0,906	22	0,771	24
2	0,928	19	0,937	9	0,795	15
3	0,902	34	0,909	20	0,765	29
4	0,934	16	0,881	36	0,776	22
5	0,905	32	0,913	18	0,779	20
6	0,872	47	0,891	30	1,000	1
7	0,947	12	0,898	25	0,774	23
8	0,874	45	0,866	48	0,746	43
9	0,890	38	0,890	32	0,812	9
10	0,881	42	0,880	38	0,759	33
11	0,909	31	0,880	37	0,752	39
12	0,942	14	0,910	19	0,763	30
13	0,854	52	0,838	55	0,720	52
14	0,844	56	0,847	53	0,701	58
15	0,883	41	0,909	21	0,721	51
16	0,887	40	0,867	46	0,706	55
17	0,857	51	0,882	35	0,709	54
18	0,913	30	0,905	24	0,822	8
19	0,875	44	0,840	54	0,703	56
20	0,937	15	0,879	39	0,781	19
21	0,934	17	0,896	28	0,808	10
22	0,899	36	0,852	49	0,755	35
23	0,918	26	0,932	10	0,793	16
24	0,943	13	0,896	26	0,769	25
25	0,959	8	0,974	3	0,826	7
26	0,905	33	0,874	43	0,750	42
27	0,846	55	0,875	42	0,751	41
28	0,970	7	0,929	11	0,797	13
29	0,869	48	0,828	56	0,702	57
30	0,889	39	0,866	47	0,743	44

31	0,848	54	0,849	51	0,729	50
32	0,823	58	0,810	58	0,730	49
33	0,989	3	0,948	6	0,796	14
34	0,925	20	0,939	8	0,806	11
35	0,862	50	1,000	1	0,847	4
36	0,991	2	0,949	5	0,826	6
37	0,872	46	0,851	50	0,719	53
38	0,925	21	0,874	44	0,759	34
39	0,915	28	0,882	34	0,755	36
40	0,901	35	0,888	33	0,782	18
41	1,000	1	0,941	7	0,848	3
42	0,914	29	0,891	31	0,761	32
43	0,921	24	0,892	29	0,769	26
44	0,868	49	0,878	40	0,742	45
45	0,894	37	0,896	27	0,768	28
46	0,948	11	0,918	15	0,776	21
47	0,983	4	0,976	2	0,849	2
48	0,977	5	0,968	4	0,827	5
49	0,923	23	0,877	41	0,736	47
50	0,877	43	0,925	13	0,786	17
51	0,918	27	0,871	45	0,732	48
52	0,827	57	0,823	57	0,737	46
53	0,953	9	0,914	17	0,768	27
54	0,976	6	0,915	16	0,762	31
55	0,923	22	0,921	14	0,754	37
56	0,919	25	0,927	12	0,753	38
57	0,931	18	0,905	23	0,802	12
58	0,851	53	0,847	52	0,752	40
Wartość minimalna	0,823		0,810		0,701	
Wartość średnia	0,910		0,895		0,771	
Współczynnik determinacji R2	0,996		0,996		0,996	

Wykres 3

Poziom efektywności poszczególnych oddziałów banku dla trzech kwartałów



oddziałów (por. wykres 3), zwłaszcza w IV kwartale (im bliżej środka wykresu znajduje się krzywa, tym mniejsza jest efektywność danego oddziału).

Niestety, ze względu na zastosowaną najprostszą (uproszczoną) metodę estymacji stochastycznego modelu granicznego kosztów (SMNK) nie można precyzyjnie wskazać źródeł powstałych nieefektywności (np. ponoszenie nadmiernych nakładów, stosowanie nieprawidłowych proporcji czynników produkcji czy produkowanie na nieoptymalnym poziomie). Z tego samego powodu uzyskane poziomy efektywności dla poszczególnych oddziałów są zazwyczaj zaniżone i ich dokładna interpretacja byłaby mało wiarygodna. Prawidłowe wnioskowanie o efektywności możliwe jest dopiero w wyniku zastosowania bardziej precyzyjnych metod estymacji, np. na gruncie bayesowskim.

Do oceny oraz porównań z innymi metodami mogą natomiast służyć otrzymane rankingi oddziałów sporządzone na podstawie uzyskanych poziomów efektywności kosztowej, ponieważ zastosowana metoda estymacji (SMNK) nie ma już wpływu na uszeregowanie otrzymanych miar efektywności. Sporządzone rankingi oddziałów należy więc uznać za wiarygodne i użyteczne w dalszej analizie i ocenie, co niewątpliwie jest istotną korzyścią badawczą.

7. Podsumowanie

Mikroekonomiczna (graniczna) funkcja kosztów, otrzymana w wyniku rozwiązania problemu minimalizacji kosztów ma szerokie zastosowanie w pomiarze efektywności kosztowej instytucji finansowych, w szczególności banków i ich oddziałów. Analiza efektywności kosztowej pozwala

na zbadanie, czy poszczególne podmioty przy danych cenach czynników produkcji ponoszą minimalny koszt całkowity wytworzenia danego poziomu produkcji, czyli umożliwia sprawdzenie, czy dana wielkość produkcji została osiągnięta przy odpowiedniej skali i strukturze tych czynników. Jeżeli przedsiębiorstwo ponosi większy koszt, niż wynika to z mikroekonomicznej (granicznej) funkcji kosztu, to przyczyną jest nieefektywność kosztowa – alokacyjna lub techniczna (Mazrec, Osiewalski 1996–1997).

Po oszacowaniu granicznej funkcji kosztów poziom efektywności kosztowej danego obiektu mierzy się przez porównanie go z obiektem wzorcowym (najbardziej efektywnym w danej próbie). Dzięki temu możliwy jest również pomiar poziomu nieefektywności, czyli można określić, ile dany obiekt utracił (np. poniósł zbyt duże nakłady lub uzyskał zbyt małe efekty w stosunku do posiadanych możliwości) wskutek nieracjonalnej działalności. Pozwala to także na wskazanie źródeł nieefektywności, np. ponoszenie nadmiernych nakładów, stosowanie nieprawidłowych proporcji czynników produkcji lub produkowanie na nieoptymalnym poziomie.

Należy podkreślić, że zaprezentowane w niniejszym artykule podejście ma silne podstawy mikroekonomiczne i statystyczne, a pomiar efektywności kosztowej następuje w wyniku rozwiązania zagadnienia optymalizacji (minimalizacji) kosztów.

Bibliografia

- Aigner D., Lovell C.A.K., Schmidt P. (1977), Formulation and Estimation of Stochastic Frontier Production Models, *Journal of Econometrics*, 6.
- Barbarski J. (2003), *Porównanie tradycyjnych i ekonometrycznych metod oceny efektywności ekonomicznej banków i ich oddziałów*, praca doktorska pod kier. J. Osiewalskiego, Akademia Ekonomiczna w Krakowie (maszynopis powielony).
- Bednarski P., Gołajewska M., Józefowska M., Nikiel E.M., Opiela T.P. (1999), *Efektywność i ryzyko sektora bankowego w Polsce*, Materiały i Studia, 96, NBP, Warszawa.
- Bednarski P., Gołajewska M., Józefowska M., Nikiel E.M., Opiela T.P. (2001), *Assessing the Efficiency of Polish Commercial Banks*, Materiały i Studia, 18, NBP, Warszawa.
- Nikiel E.M., Opiela T.P. (2002), Customer Type of Bank Efficiency in Poland: Implication for Emerging Market Banking, *Contemporary Economic Policy*, 20 (3).
- Berger A.N., Hancock D., Humphrey D.B. (1993), Bank Efficiency Derived from the Profit Function, *Journal of Banking and Finance*, 17.
- Berger A.N., Hunter W.C., Timme S.G. (1993), The Efficiency of Financial Institutions: A Review and Preview of Research Past, Present and Future, *Journal of Banking and Finance*, 17.
- Christensen L.R., Jorgenson D., Lau L. (1971), Conjugate Duality and the Transcendental Logarithmic Production Function, *Econometrica*, 39.
- English M., Grosskopf S., Hayes K., Yaiswarng S. (1993), Output Allocative and Technical Efficiency of Banks, *Journal of Banking and Finance*, 17.
- Farrell M. (1957), The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society, Series A* (120).
- Gabrielsen A. (1975), *On Estimating Efficient Productions*, Working Paper, A – 35, Chr. Michelsen Institute, Department of Humanities and Social Sciences, Bergen.

- Gospodarowicz M. (2000), *Procedury, analizy i oceny banków komercyjnych*, Materiały i Studia, 113, NBP, Warszawa.
- Greene W.H. (1990), A Gamma – Distributed Stochastic Frontier Model, *Journal of Econometrics*, 46.
- Koop R., Diewert W.E. (1982), The Decomposition of Frontier Cost Function Deviations into Measures of Technical and Allocative Inefficiency, *Journal of Econometrics*, 19.
- Kopp R., Mullahy J. (1989), *Moment – Based Estimation and Testing of Stochastic Frontier Models*, Discussion Papers, 89 – 100, Resources for the Future, Washington.
- Marzec J. (1998), *Produkty i czynniki produkcji w badaniach efektywności kosztowej banków*, Prace Naukowe AE we Wrocławiu.
- Marzec J. (1999), Produkty, czynniki produkcji i funkcja kosztów w badaniach efektywności kosztowej banków, *Ekonomista*, 3.
- Marzec J. (2000), *Ekonometryczna analiza efektywności kosztów w bankach komercyjnych*, praca doktorska pod kier. J. Osiewalskiego, Akademia Ekonomiczna w Krakowie (maszynopis powielony).
- Marzec J. (2002), *Krótkookresowa analiza technologii i efektywności kosztowej oddziałów banku – praca jako czynnik stały*, w: A. Welfe (red.), *Metody ilościowe w naukach ekonomicznych*, Zeszyty Naukowe SGH, Warszawa.
- Marzec J., Osiewalski J. (1996–1997), *Pomiar efektywności kosztowej banków: zarys metodologii*, *Folia Oeconomica Cracoviensia*, t. 39–40, PAN – Oddział w Krakowie.
- Meeusen W., van den Broeck J. (1977), Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error, *International Economic Review*, 8.
- Osiewalski J. (2001), *Ekonometria bayesowska w zastosowaniach*, Wydawnictwo AE w Krakowie.
- Osiewalski J., Marzec J. (1998a), *Bayesian Analysis of Cost Efficiency with an Application to Bank Branches*, w: E. Miklaszewska (red.), *Global Trends and Changes in European Banking*, Uniwersytet Jagielloński, Kraków.
- Osiewalski J., Marzec J. (1998b), *Bayesowska analiza efektywności kosztowej oddziałów banku: założenia i wyniki [w:] Prognozowanie w zarządzaniu firmą*, Prace Naukowe AE we Wrocławiu, 808.
- Osiewalski J., Wróbel-Rotter R. (1999), Estymacja granicznych funkcji produkcji i wskaźników efektywności technicznej na podstawie danych przekrojowych, *Przegląd Statystyczny*, 1.
- Osiewalski J., Wróbel-Rotter R. (2002), Bayesowski model efektów losowych w analizie efektywności kosztowej, *Przegląd Statystyczny*, 2.
- Pawłowska M. (2005), *Poziom konkurencji i efektywność a koncentracja na polskim rynku bankowym*, Materiały i Studia, 192, NBP, Warszawa.
- Richmond J. (1974), Estimating the Efficiency of Production, *International Economic Review*, 15.
- Rogers K.E. (1998), Nontraditional Activities and the Efficiency of US Commercial Banks, *Journal of Banking and Finance*, 22.
- Rogowski G. (1998), *Metody analizy i oceny działalności banku na potrzeby zarządzania strategicznego*, Wydawnictwo Wyższej Szkoły Bankowej, Poznań.
- Sealey C.W., Lindley J.T. (1977), Inputs, Outputs and a Theory of Productions and Cost at Depository Financial Institutions, *Journal of Finance*, 32.
- Van den Broeck J., Koop G., Osiewalski J., Steel M.F.J. (1994), Stochastic Frontier Models: A Bayesian Perspective, *Journal of Econometrics*, 61.
- Varian H.L. (1999), *Mikroekonomia*, PWN, Warszawa 1999.

- Winsten C.D. (1957), Discussion on Mr. Farrell's Paper, *Journal of the Royal Statistical Society, Series A* (120).
- Wróbel-Rotter R. (2001), Giętkie formy funkcyjne w empirycznej analizie kosztu. Podejście Diewerta i wnioskowanie bayesowskie, *Ekonomista* (5).

Podziękowania

Autor pragnie w tym miejscu złożyć Jackowi Osiewalskiemu i Jerzemu Marcowi wyrazy serdecznego podziękowania za bardzo cenne uwagi merytoryczne oraz liczne i owocne dyskusje naukowe w trakcie przygotowywania pracy doktorskiej, będące przyczynkiem do powstania niniejszej publikacji.

Econometric measurement of financial institutions' economic efficiency. Stochastic cost frontier model

Abstract

The main goal of the article is the presentation of the stochastic cost frontier model as one of the ways of measuring financial institutions' economic efficiency and the application of this model to the measurement of cost efficiency of the branches of one of the Polish banks. The analysis of cost efficiency allows to examine whether, at certain prices of production factors and at a certain level of production, individual subjects incur the minimum total production cost.

This model was constructed on the basis of two variable production factors (the value of acquired deposits – financial capital, and the value of remuneration – human capital), one fixed factor (the area of office premises – real capital) and one product (the value of granted credits).

The empirical results of calculations provide indexes for the cost efficiency level and ranks of the tested branch group in particular quarters of the year. On account of an oversimplified method of estimation (Corrected Ordinary Least Squares), the derived indexes of efficiency are too low and insufficiently precise for further analysis. However, the ranks made on their basis are reliable and could be used for further evaluation and comparisons, which is their essential merit.

Keywords: economic efficiency of banks, Stochastic Cost Frontier Model, analysis and measuring of financial institutions, cost efficiency, optimisation (minimization) of costs

