

Uniwersytet Warszawski  
Wydział Nauk Ekonomicznych

Krzysztof Szafrński

Nr albumu: 386 118

**Determinanty liczby medali zdobywanych na  
igrzyskach olimpijskich**

Warszawa, lipiec 2022

## **Wstęp**

Igrzyska olimpijskie to najstarsza i zarazem największa impreza sportowa na świecie. Co 4 lata ponad 10 tysięcy sportowców z ponad 200 państw staje do rywalizacji o olimpijskie medale. Interesujące jest pytanie, co determinuje sukces kraju na igrzyskach. Intuicyjnie wydaje się, że powinien być on powiązany z liczbą ludności, która odzwierciedlałaby liczbę utalentowanych sportowców, a także z poziomem rozwoju gospodarczego. Na przykładzie samej Polski nie wydaje się to jednak tak oczywiste, biorąc pod uwagę, że osiągała ona zdecydowanie lepsze wyniki przed transformacją gospodarczą.

Celem niniejszej pracy jest właśnie zbadanie, które czynniki wpływają na liczbę medali zdobywanych na igrzyskach olimpijskich. Do analizy tego tematu zdecydowano się wykorzystać model Tobit. Praca w dużej mierze bazuje na artykule Scelles'a i in. (2020), jednak wykorzystuje bardziej aktualną próbę z lat 2000-2020.

Pracę rozpoczęto od krótkiego przeglądu literatury. Następnie opisano wykorzystane zmienne, a także uzasadniono wybór metody badawczej. Na końcu przedstawiono wyniki przeprowadzonej analizy.

## Przegląd literatury

Jednym z najczęściej cytowanych badań w tym temacie jest praca Bernarda i Busse (2004). Autorzy skupiają się na roli populacji oraz rozwoju gospodarczego w determinowaniu olimpijskiego sukcesu. Zauważyli, że nieprawdziwa jest prosta hipoteza, sugerująca, że udział państw w zdobyciach medalowych powinien być proporcjonalny do ich populacji. Wyniki wykorzystanego przez nich modelu Tobit na danych z lat 1960-1996 wskazywały, że obok ludności najistotniejszą determinantą jest PKB per capita. Ponadto średnio więcej medali zdobywają również gospodarze igrzysk, a także państwa z bloku wschodniego.

Moosa i Smith (2004) rozszerzyli listę potencjalnych zmiennych objaśniających o 10 różnych wskaźników społeczno-ekonomicznych oraz o wielkość reprezentacji na igrzyskach. Oprócz ludności i PKB istotne i pozytywnie związane z liczbą zdobywanych medali były również wydatki na zdrowie, a także wielkość reprezentacji. Autorzy jednak przeprowadzili swoją analizę wyłącznie na danych z igrzysk w Sydney.

Wielkość reprezentacji była również najważniejszym predyktorem w badaniu Vagenasa i Vlachokyriakou (2012), którzy z kolei analizowali igrzyska w Atenach. Ponownie istotne były też wydatki na zdrowie, a także stopa wzrostu PKB i bezrobocie. Autorzy dowodzą, że tradycyjny model skupiający się tylko na znaczeniu PKB i populacji jest niewystarczający do wyjaśnienia sukcesu na igrzyskach. Wydaje się jednak, że wniosków z tego badania nie należy zbyt uogólniać, ponieważ autorzy wykorzystali dane z tylko jednych igrzysk, a także ograniczyli próbę do państw, które zdobyły co najmniej 1 medal.

Andreff (2013) wzorując się na pracy Bernarda i Busse (2004) wykorzystał model Tobit na danych z lat 1976-2004. Do analizowanych wcześniej predyktorów dodał dwie zmienne dyskretne wskazujące na reżim polityczny panujący w danym państwie, a także region geograficzny, mający odzwierciedlać różnice kulturowe. Obie zmienne okazały się istotne. W szczególności dobrze na tle innych państw wypadały państwa komunistyczne i postkomunistyczne. Model również zadowalająco prognozował dorobek medalowy na igrzyskach w Pekinie. Faktyczna liczba zdobytych w medali w 70% przypadków mieściła się w prognozowanym przedziale ufności.

Scelles i in. (2020) przeprowadzili podobne badanie na nowszej próbie z lat 1992-2016, jednak oprócz modelu Tobit wykorzystali również model Hurdle. Nie było znacznych różnic między modelami w kontekście istotności poszczególnych zmiennych, natomiast model Hurdle dawał nieznacznie lepsze prognozy na igrzyska w Rio de Janeiro.

## Opis zmiennych i metoda badawcza

Do przeprowadzenia analizy zebrano dane z sześciu letnich igrzysk olimpijskich z lat 2000-2020 ze strony <https://www.wikipedia.org/>. Natomiast dane na temat wskaźników społeczno-ekonomicznych zaczerpnięto z Banku Światowego. Poniżej przedstawiono opis wykorzystanych zmiennych:

- *medale* - całkowita liczba medali zdobytych przez dane państwo
- *gospodarz* - zmienna binarna, wskazująca na gospodarza igrzysk
- *sportowcy* - liczba atletów reprezentujących dane państwo na igrzyskach
- *pkb\_pc* - PKB per capita w cenach bieżących w dolarach amerykańskich
- *populacja* - całkowita ludność
- *zdrowie* - całkowite wydatki na zdrowie jako procent PKB
- *region* - zmienna dyskretna wskazująca na region, w którym położone jest państwo. Wzorowana jest na poprzednich badaniach (Andreff, 2013; Scelles i in., 2020). Przyjmuje 9 poziomów: *AFS* - Afryka Subsaharyjska, *AFN* - Afryka północna, *MNE* - Bliski Wschód, *ASI* - reszta Azji, *OCE* - Australia i Oceania, *NAM* - Ameryka Północna i Karaiby, *LSA* - Ameryka Łacińska, *WEU* - Europa Zachodnia, *EAST* - Europa Wschodnia.

Podobnie jak w poprzednich badaniach (Andreff, 2013; Scelles i in., 2020) zmienne *pkb\_pc*, *populacja* i *zdrowie* są opóźnione o 4 lata względem igrzysk. Na przykład dla igrzysk w 2000 roku podane jest PKB per capita danego kraju z 1996 roku. Zakłada to, że państwo potrzebuje czasu, żeby skorzystać z dostępnych zasobów i przygotować się do igrzysk.

Biorąc pod uwagę liczebnościowy charakter zmiennej zależnej oraz duży udział obserwacji zerowych, wydaje się, że odpowiednim modelem do analizy tego zjawiska byłby *zero-inflated Poisson* lub *zero-inflated negative binomial*. Jednak model Poissona zakłada równość średniej oraz wariancji rozkładu, a warunek ten zdecydowanie nie jest spełniony w przypadku liczby zdobytych medali (średnia = 4,75; wariancja = 176). Model *negative binomial* już nie jest oparty na tym założeniu, jednak występuje tu inny problem związany z modelami *zero-inflated*. Zakładają one, że obserwacje zerowe mają dwa źródła: “próbkowe” i “strukturalne” (Hu i in., 2011). Pierwsze dotyczy obserwacji, które przyjmują zero w sposób naturalny, natomiast drugie obejmuje obserwacje, których dane zjawisko w ogóle nie

dotyczy. Na przykład w kontekście liczby zdobytych medali zera “strukturalne” występowałyby w przypadku państw, które nie wzięły udziału w igrzyskach. Jednak w analizowanej próbie uwzględniono tylko państwa uczestniczące w igrzyskach, dlatego modele *zero-inflated* nie wydają się odpowiednie.

Z powyższych powodów do analizy zdecydowano się wykorzystać model Tobit, który był też najczęściej stosowany w literaturze. Poniżej przedstawiono postać modelu:

$$\begin{aligned} medale_{i,t}^* = & \beta_0 + \beta_1 \ln(pkb\_pc)_{i,t-4} + \beta_2 \ln(populacja)_{i,t-4} + \beta_3 zdrowie_{i,t-4} \\ & + \beta_4 gospodarz_{i,t} + \beta_5 \ln(sportowcy)_{i,t} + \sum_n \beta_n region_i + u_i + \varepsilon_{i,t} \end{aligned}$$

$$medale_{i,t} = medale_{i,t}^*, \text{ jeśli } medale_{i,t}^* > 0$$

$$medale_{i,t} = 0, \text{ jeśli } medale_{i,t}^* \leq 0$$

Na podstawie literatury należy się spodziewać, że zmienne *pkb\_pc*, *populacja*, *gospodarz* oraz *sportowcy* będą istotne i dodatnio powiązane ze zmienną zależną. Przy zmiennej *zdrowie* również oczekuje się dodatniego parametru, choć nie była ona zbyt często analizowana w innych badaniach. Natomiast w przypadku zmiennej *region* najistotniejszym poziomem była zazwyczaj Ameryka Północna.

## Wyniki modelu

Poniżej przedstawiono wyniki dwóch oszacowanych modeli Tobit - z efektami losowymi oraz z efektami stałymi.

Tabela 1. Wyniki modelu Tobit z efektami losowymi (Model 1) oraz stałymi (Model 2)

	Model 1	Model 2
(Intercept)	-97.34 *** (5.76)	4.26 (3.15)
log(pkb_pc)	0.78 * (0.36)	-1.39 (1.36)
log(populacja)	3.77 *** (0.28)	5.19 *** (0.12)
zdrowie	0.86 *** (0.14)	0.13 (2.45)
log(sportowcy)	6.52 *** (0.45)	2.63 ** (0.89)
gospodarz1	12.58 *** (2.01)	9.04 *** (1.14)
regionAFN	-4.52 ** (1.53)	
regionAFS	2.84 (1.67)	
regionASI	4.51 *** (1.30)	
regionEAST	9.74 *** (1.24)	
regionLSA	0.65 (1.25)	
regionMNE	11.10 *** (1.31)	
regionNAM	17.62 *** (1.17)	
regionOCE	8.68 *** (1.69)	
igrzyska2004	-0.65 (0.71)	-4.27 * (1.95)
igrzyska2008	-0.57 (0.71)	-7.99 *** (2.00)
igrzyska2012	-1.60 * (0.78)	-11.60 *** (1.90)
igrzyska2016	-2.66 *** (0.80)	-16.03 *** (2.18)
igrzyska2020	-1.33 (0.80)	-19.28 *** (1.57)
sd.nu	4.65 *** (0.16)	
sd.eta	10.59 *** (0.23)	
Num. obs.	1095	906
Log Likelihood	-1680.75	
R <sup>2</sup>		0.08
Adj. R <sup>2</sup>		0.07
*** p < 0.001; ** p < 0.01; * p < 0.05		

Źródło: opracowanie własne.

W przypadku modelu z efektami losowymi rzeczywiście zmienne *pkb\_pc*, *populacja*, *gospodarz*, *sportowcy* i *zdrowie* są istotne i pozytywnie związane z liczbą zdobywanych na igrzyskach medali. Co ciekawe, większość regionów osiąga lepsze wyniki niż wysoko rozwinięta Europa Zachodnia, która jest poziomem bazowym. Wyjątkiem są oba regiony Afryki oraz Ameryka Łacińska. Jednak przeprowadzony test Hausmana wskazuje, że istnieje korelacja między efektem indywidualnym a zmiennymi objaśniającymi, co prowadzi do niezgodności estymatora efektów losowych.

Powinno się zatem skupić na modelu z efektami stałymi, który prowadzi do nieco innych wyników. Przede wszystkim nieistotna okazała się zmienna *pkb\_pc*, co jest dość zaskakujące w kontekście poprzednich badań. Ponadto nieistotne są również wydatki na zdrowie. Natomiast oszacowania przy zmiennych *populacja*, *gospodarz* i *sportowcy* nie zmieniły się znacząco względem pierwszego modelu. Niestety w modelu nie da się oszacować wpływu zmiennych wskazujących na region, ponieważ są one stałe w czasie.

## **Podsumowanie**

Celem niniejszej pracy była odpowiedź na pytanie, co determinuje sukces państw na igrzyskach olimpijskich. Wyniki przeprowadzonego modelu Tobit sugerują, że lepsze wyniki osiągają kraje z większą populacją oraz te, które wystawiają na igrzyska większą reprezentację. Ponadto istotny był także efekt gospodarza, czyli więcej medali zdobywają również państwa organizujące igrzyska. Jednak nieistotne okazało się PKB per capita, co jest sprzeczne z wynikami większości przeprowadzonych wcześniej badań. Ciężko powiedzieć czym jest to spowodowane. Wydaje się, że nie powinna być to kwestia przyjętej metody badawczej, ponieważ model Tobit był często wykorzystywany we wcześniejszych pracach. Główną część prób z poprzednich badań stanowiły igrzyska sprzed 2000 roku, za to niniejsza praca wykorzystuje wyłącznie obserwacje z lat 2000-2020. Możliwe zatem, że z jakiegoś powodu w XXI wieku PKB straciło na znaczeniu, w kontekście determinowania sukcesu na igrzyskach.



## Literatura

- Andreff, W. (2013). Economic development as major determinant of Olympic medal wins: predicting performances of Russian and Chinese teams at Sochi Games. *International Journal of Economic Policy in Emerging Economies*, 6(4), 314-340.
- Bernard, A. B., & Busse, M. R. (2004). Who wins the Olympic Games: Economic resources and medal totals. *Review of economics and statistics*, 86(1), 413-417.
- Hu, M. C., Pavlicova, M., & Nunes, E. V. (2011). Zero-inflated and hurdle models of count data with extra zeros: examples from an HIV-risk reduction intervention trial. *The American journal of drug and alcohol abuse*, 37(5), 367-375.
- Moosa, I. A., & Smith, L. (2004). Economic development indicators as determinants of medal winning at the Sydney Olympics: an extreme bounds analysis. *Australian economic papers*, 43(3), 288-301.
- Scelles, N., Andreff, W., Bonnal, L., Andreff, M., & Favard, P. (2020). Forecasting national medal totals at the Summer Olympic Games reconsidered. *Social Science Quarterly*, 101(2), 697-711.
- Vagenas, G., & Vlachokyriakou, E. (2012). Olympic medals and demo-economic factors: Novel predictors, the ex-host effect, the exact role of team size, and the “population-GDP” model revisited. *Sport Management Review*, 15(2), 211-217.