

PROBLEM ELIMINACJI TRENDÓW PODCZAS OCENY RYZYKA PRODUKCYJNEGO PODSTAWOWYCH ROŚLIN UPRAWNYCH¹

Paweł Kobus

Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych
Szkoły Głównej Gospodarstwa Wiejskiego w Warszawie
Kierownik katedry: dr hab. Maria Parlińska, prof. SGGW

Słowa kluczowe: plony roślin uprawnych, prognozowanie, trend, ryzyko
Key words: crops yields, prediction, risk

S y n o p s i s. Na podstawie danych dotyczących plonowania roślin uprawnych z wielu lat można potwierdzić występowanie systematycznych zmian w wysokości plonu na poziomie krajowym. W ocenie wielkości ryzyka produkcyjnego ogólną zmienność plonowania pomniejsza się o systematyczne zmiany wynikające z trendów w plonowaniu poszczególnych roślin uprawnych. W pracy podjęto próbę oceny zasadności takiego postępowania. W tym celu wyznaczono trendy plonowania podstawowych roślin uprawnych w Polsce i porównano jakość predykcji plonów opartych na wyznaczonych trendach oraz jakość predykcji plonów bazujących na średniej arytmetycznej. Jako miarę jakości predykcji przyjęto pierwiastek ze średniego kwadratu błędu predykcji. Badania wykazały, że stosowanie funkcji trendu do predykcji plonowania prowadzi przeciętnie do pogorszenia jakości predykcji o około 27%.

WPROWADZENIE

Powszechnie uważa się, że plony roślin uprawnych systematycznie wzrastają. Przypisuje się to z dwóm głównym przyczynom: postępowi biologicznemu i postępowi technologicznemu [Wicki 2010, s. 168]. Wysokość plonów roślin uprawnych analizowanych w dłuższym czasie zagregowanych na poziomie kraju potwierdza występowanie takich trendów [Kobus, Wicki 2010, s. 109-115, Peltonen-Sainio i in. 2009, s. 85-90]. Powstaje pytanie, jak postępować z trendem podczas oceny wielkości ryzyka produkcyjnego w produkcji roślinnej. Ogólną zmienność plonowania można podzielić na czystą zmienność losową i zmienność wynikającą z systematycznych zmian w przeciętnym plonowaniu kojarzoną z funkcją trendu. Nieuwzględnienie występującego trendu prowadzi do zawyżenia oceny wielkości ryzyka [Marra, Schurle 1994, 69-77]. Dlatego niektórzy badacze zalecają usunięcie trendu z danych przed określeniem wielkości ryzyka [Moss 2010, s. 44-45].

Pozostaje jednak kilka otwartych kwestii: po pierwsze, na ile stabilny jest trend w długim okresie. Brak takiej stabilności oznaczałby konieczność szacowania trendu tylko

¹ Badania zrealizowano w ramach projektu Narodowego Centrum Nauki nr 3916/B/H03/2011/40 pt. *Metody pomiaru i ograniczania ryzyka w produkcji rolniczej w Polsce w warunkach przemian instytucjonalnych i zmian klimatycznych.*

na podstawie kilku lub kilkunastu ostatnich lat. Kolejną kwestią jest przeniesienie trendu z poziomu kraju lub regionu na poziom gospodarstwa rolnego. Nawet jeżeli na poziomie kraju można zaobserwować wyraźny trend, to w pojedynczych gospodarstwach może to wyglądać całkiem inaczej. I ostatnia kwestia: sprawdzenie, czy prognozowanie przyszłych plonów na podstawie funkcji trendu daje trafniejsze prognozy niż użycie zwykłej średniej i tym samym potwierdza realną wartość funkcji trendu w wyjaśnianiu zmienności plonów.

Podstawowym celem pracy jest określenie zasadności eliminacji trendu przed wyznaczeniem wielkości ryzyka produkcyjnego w produkcji roślinnej na poziomie gospodarstwa rolnego oraz sprawdzenie, czy wskazane jest używanie funkcji trendu do prognozowania przyszłych plonów.

DANE I METODY

W pracy wykorzystano dane polskiego FADN z lat 2004-2012. Ponieważ podstawowy cel pracy był związany z oszacowaniem trendów w plonowaniu roślin uprawnych w Polsce na poziomie gospodarstwa rolnego, wydzielono dla każdej rośliny oddzielnie próbę gospodarstw, dla których było dostępnych co najmniej 5 obserwacji plonów. W rezultacie uzyskano następujące rozmiary prób (liczebność) dla poszczególnych roślin:

- pszenica ozima – 5242,
- pszenica jara – 1356,
- żyto – 3350,
- jęczmień jary – 4686,
- owies – 1668,
- pszenżyto ozime – 5637,
- mieszanki zbożowe na ziarno – 4976,
- kukurydza ziarno – 590,
- ziemniaki – 5773,
- buraki cukrowe – 1527,
- rzepak i rzepik – 1971.

Z uwagi na stosunkowo krótkie szeregi czasowe, tzn. maksymalnie 9 lat dla gospodarstwa, ograniczono się do modeli regresji z liniową funkcją trendu:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad t = 1, \dots, n \quad (1)$$

gdzie: Y_t – plon w roku t , ε_t – składnik losowy w roku t , n – liczba obserwacji dostępnych dla wybranego gospodarstwa.

Należy zwrócić uwagę, że dla poszczególnych gospodarstw były dostępne różne liczby obserwacji od 1 do 9. W pracy ograniczono się do gospodarstw, dla których było dostępnych co najmniej 5 obserwacji. Model (1) został oszacowany oddzielnie dla każdego gospodarstwa, przy czym do estymacji użyto tylko $n-1$ pierwszych obserwacji. Dzięki takiemu postępowaniu możliwe było określenie faktycznego błędu predykcji dla ostatniej obserwacji:

$$BP_T = y_n - (\hat{\beta}_0 + \hat{\beta}_1 n) \quad (2)$$

W przypadku użycia wszystkich dostępnych obserwacji do oszacowania modelu wartość obliczona ze wzoru (2) nie byłaby błędem predykcji, lecz błędem dopasowania.

Rozróżnienie to traci na znaczeniu w przypadku dużej liczby obserwacji, jednak dla małej liczebności próby różnica ta jest bardzo istotna.

Alternatywnym modelem do modelu (1) był model zakładający brak trendu:

$$Y_t = \mu + \varepsilon_t, \quad t = 1, \dots, n \tag{3}$$

gdzie: μ – wartość oczekiwana dla plonu w gospodarstwie.

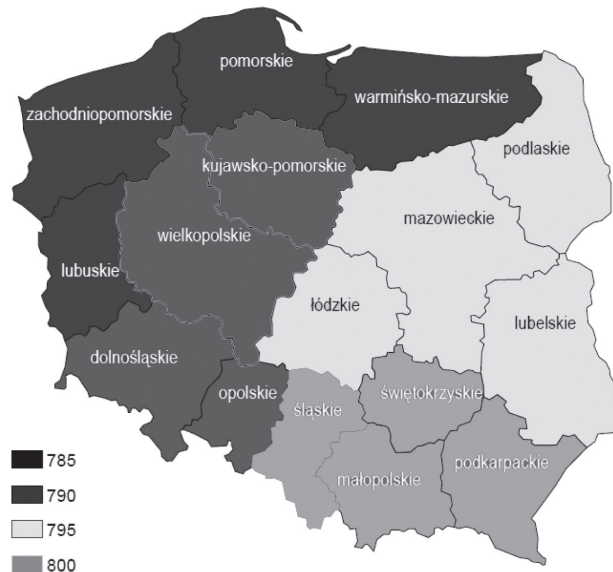
W przypadku modelu (3) oszacowanie modelu sprowadza się do obliczenia średniej arytmetycznej \bar{Y} , również w tym przypadku do oceny modelu użyto $n - 1$ obserwacji. Dla modelu (3) błąd predykcji przyjmuje następującą postać:

$$BP_s = y_n - \bar{Y} \tag{4}$$

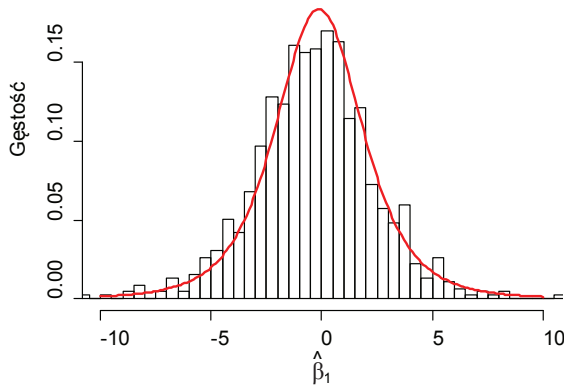
WYNIKI BADAŃ

W pracy poddano analizie plonowanie podstawowych roślin uprawnych w Polsce, tzn. roślin charakteryzujących się największym udziałem w strukturze zasiewów. Zgodnie z danymi z *Rocznika statycznego rolnictwa* [GUS 2013] były to: pszenica ozima, pszenica jara, żyto, jęczmień jary, owies, pszenżyto ozime, mieszanki zbożowe na ziarno, kukurydza na ziarno, ziemniaki, buraki cukrowe, rzepak i rzepik. Uprawy te łącznie stanowiły ponad 82% ogólnej powierzchni zasiewów w 2012 r. Dla każdej z wymienionych roślin wykonano oddzielne obliczenia z podziałem na 4 regiony kraju wydzielone zgodnie z rysunkiem 1.

Modele regresji liniowej oszacowano oddzielnie dla każdego gospodarstwa, dzięki temu dla poszczególnych regionów i roślin dysponowano dużą liczbą oszacowań paramet-



Rysunek 1. Podział Polski na regiony zgodnie z metodologią FADN
 Źródło: [Goraj i inni 2012, s. 5].



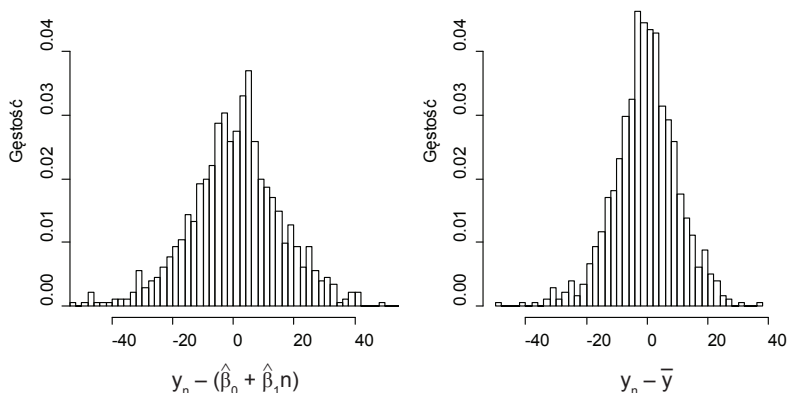
Rysunek 2. Rozkład ocen współczynnika kierunkowego funkcji trendu dla plonów pszenicy ozimej w gospodarstwach z regionu 790

Źródło: obliczenia własne na podstawie danych FADN.

tru β_1 . Na rysunku 2. w postaci histogramu przedstawiono rozkład ocen parametru β_1 dla plonów pszenicy ozimej w regionie 790 na podstawie obserwacji dla 2307 gospodarstw.

Rozkład z rysunku 2. wskazuje, że w większości przypadków oceny parametru β_1 są bliskie zera. Formalnie, przy założeniu, że błąd ε_i z modelu (1) podlega rozkładowi normalnemu $N(0, \sigma^2)$, iloraz $\frac{\hat{\beta}_1 - \beta_1}{S_{\beta_1}}$ podlega rozkładowi t-Studenta. Konsekwentnie

rozkład $\hat{\beta}_1$ powinien być skoncentrowany wokół β_1 i zbliżony do przeskalowanego rozkładu t-Studenta. Linia ciągła z rysunku 2. jest funkcją gęstości takiego właśnie przeskalowanego rozkładu t-Studenta o stopniach swobody równych 4. Umiejscowienie osi symetrii tego rozkładu w okolicy zera wskazuje na to, że prawdziwa wartość β_1 jest bliska zera. Powyższe fakty wskazują na brak systematycznych zmian w plonowaniu pszenicy ozimej w latach 2004-2012 w regionie 790. W konsekwencji używanie funkcji trendu



Rysunek 3. Rozkład błędów predykcji dla modelu z funkcją trendu i modelu zawierającego tylko średnią

Źródło: obliczenia własne na podstawie danych FADN.

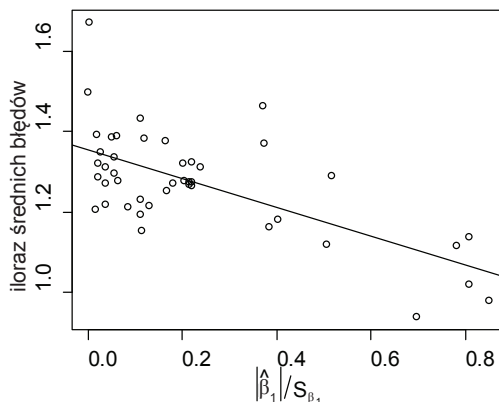
do prognozowania plonów będzie prowadziło do powiększenia błędu prognozy zamiast spodziewanego zmniejszenia.

Chociaż dla obydwu modeli błędy prognozy są duże, to jednak rozkład błędów dla modelu bez trendu jest wyraźnie bardziej skoncentrowany wokół zera. Przyjmując jako średni błąd pierwiastek ze średniego kwadratu błędu predykcji, otrzymuje się dla modelu z trendem 13,4 dt, podczas gdy dla modelu bez trendu – 10,5 dt, co oznacza, że użycie modelu z trendem pogarsza jakość predykcji o około 28%. Warto przy tej okazji zwrócić jeszcze raz uwagę na znaczenie metody oceny błędu predykcji. W przeprowadzonej analizie błąd predykcji był wyznaczany dla obserwacji, które nie były użyte podczas estymacji modeli. Gdyby obserwacje te były użyte do estymacji modeli, średnie błędy predykcji dla pszenicy ozimej w regionie 790 wyniosłyby 6,9 dt dla modelu z trendem i 9,0 dla modelu bez trendu. Jednak takie postępowanie w przypadku małych prób prowadzi do zafałszowania wyników i całkowicie błędnych wniosków. W tabeli 2. podane zostały sumaryczne wyniki przeprowadzonej analizy dla podstawowych roślin uprawnych z podziałem na regiony.

Dane w tabeli 2. wskazują, że w przypadku zbóż przeciętne wartości ocen parametru β_1 są bliskie zera w porównaniu z odchyleniami standardowymi. Wskazywać to może, że nie tylko w przypadku pszenicy, ale również pozostałych zbóż nie nastąpiła w ostatnich latach poprawa plonowania. W przypadku kukurydzy i buraków cukrowych obserwuje się dodatnie wartości współczynnika kierunkowego zbliżone wielkością do ich odchyłeń standardowych, co potwierdza systematyczne zwiększanie plonów dla tych upraw. Z kolei w przypadku rzepaku średnie wartości współczynników kierunkowych wskazują na obniżenie plonowania w latach 2004-2012.

W przypadku przeciętnych wartości współczynników kierunkowych bliskich zera korzystanie z funkcji trendu do prognozowania przyszłych plonów powinno prowadzić do zwiększenia średnich błędów predykcji. Wielkości z tabeli 2. potwierdzają tę tezę, w przypadku zbóż użycie funkcji trendu prowadzi do zwiększenia średnich błędów predykcji przeciętnie o 32% w porównaniu z prognozami opartymi na średniej arytmetycznej. W odniesieniu do pozostałych upraw tylko w przypadku kukurydzy i buraków cukrowych w regionie 790 zaobserwowano niewielką poprawę jakości predykcji, odpowiednio o 6,4% i 2,2%. Powstaje w tej sytuacji pytanie: jak duże muszą być wartości współczynników kierunkowych, aby uzyskać poprawę jakości predykcji w porównaniu ze zwykłą średnią.

Z rysunku 4. wynika, że jakość predykcji modelu bazującego na trendzie zaczyna przewyższać jakość predykcji zwykłej średniej arytmetycznej, dopiero gdy przeciętne wartości $\hat{\beta}_1$ są zbliżone do wartości odchylenia standardowego $\hat{\beta}_1$.



Rysunek 4. Zależność stosunku średniego błędu predykcji modelu z trendem do średniego błędu modelu bez trendu od standaryzowanej bezwzględnej wartości średniej $\hat{\beta}_1$

Źródło: obliczenia własne na podstawie danych FADN.

Tabela 2. Rozkłady ocen współczynników kierunkowych funkcji trendu oraz przeciętne błędy predykcji dla podstawowych roślin uprawnych w Polsce

Uprawa	Rodzaj parametru	Wielkość w regionie			
		785	790	795	800
Pszenica ozima	średnie $\hat{\beta}_1$	-0,281	-0,129	-0,439	-0,323
	odchylenie standardowe $\hat{\beta}_1$	2,148	2,038	1,997	1,929
	liczba gospodarstw	723	2307	1387	825
	średni błąd predykcji dla modelu z trendem	12,767	13,452	13,139	11,405
	średni błąd predykcji dla modelu bez trendu	10,511	10,539	9,939	9,103
Pszenica jara	średnie $\hat{\beta}_1$	-0,131	-0,005	-0,228	-0,388
	odchylenie standardowe $\hat{\beta}_1$	2,356	1,884	2,039	1,809
	liczba gospodarstw	242	346	540	228
	średni błąd predykcji dla modelu z trendem	12,882	15,852	12,034	11,361
	średni błąd predykcji dla modelu bez trendu	9,943	9,474	8,407	8,959
Żyto	średnie $\hat{\beta}_1$	0,404	-0,034	-0,076	-0,054
	odchylenie standardowe $\hat{\beta}_1$	1,994	1,687	1,382	1,448
	liczba gospodarstw	450	1258	1448	194
	średni błąd predykcji dla modelu z trendem	13,349	11,580	9,303	9,808
	średni błąd predykcji dla modelu bez trendu	10,469	9,008	6,962	7,720
Jęczmień jary	średnie $\hat{\beta}_1$	-0,041	-0,105	-0,370	-0,379
	odchylenie standardowe $\hat{\beta}_1$	2,110	2,084	1,844	1,713
	liczba gospodarstw	673	2149	1271	593
	średni błąd predykcji dla modelu z trendem	12,979	13,242	11,256	11,394
	średni błąd predykcji dla modelu bez trendu	9,326	9,558	8,529	8,956
Owies	średnie $\hat{\beta}_1$	-0,074	-0,001	-0,035	-0,182
	odchylenie standardowe $\hat{\beta}_1$	2,064	1,877	1,666	1,618
	liczba gospodarstw	326	409	635	298
	średni błąd predykcji dla modelu z trendem	12,225	12,374	9,666	10,077
	średni błąd predykcji dla modelu bez trendu	9,327	8,268	7,321	8,199
Pszennyto ozime	średnie $\hat{\beta}_1$	0,031	-0,349	-0,395	-0,417
	odchylenie standardowe $\hat{\beta}_1$	1,885	1,922	1,803	1,737
	liczba gospodarstw	740	2271	2260	366
	średni błąd predykcji dla modelu z trendem	11,519	12,272	11,467	11,488
	średni błąd predykcji dla modelu bez trendu	9,551	9,656	9,064	8,766
Mieszanka zbożowe na ziarno	średnie $\hat{\beta}_1$	-0,102	-0,322	-0,044	-0,397
	odchylenie standardowe $\hat{\beta}_1$	1,624	2,010	1,536	1,748
	liczba gospodarstw	538	1636	2349	453
	średni błąd predykcji dla modelu z trendem	10,768	12,482	9,121	10,111
	średni błąd predykcji dla modelu bez trendu	7,728	9,099	6,784	7,915
Kukurydza ziarno	średnie $\hat{\beta}_1$	3,225	3,061	1,429	1,257
	odchylenie standardowe $\hat{\beta}_1$	3,991	4,392	12,113	3,274
	liczba gospodarstw	25	308	103	154
	średni błąd predykcji dla modelu z trendem	24,179	24,602	34,762	23,986
	średni błąd predykcji dla modelu bez trendu	23,733	26,293	25,163	20,638

Tabela 2. Cd.

Ziemniaki	średnie $\hat{\beta}_i$	1,874	1,944	1,017	-0,425
	odchylenie standardowe $\hat{\beta}_i$	14,031	14,786	11,372	11,528
	liczba gospodarstw	798	1727	2457	791
	średni błąd predykcji dla modelu z trendem	70,445	81,410	67,279	76,942
	średni błąd predykcji dla modelu bez trendu	62,331	69,815	55,720	63,135
Buraki cukrowe	średnie $\hat{\beta}_i$	9,910	17,095	14,256	9,818
	odchylenie standardowe $\hat{\beta}_i$	19,616	20,103	17,630	18,968
	liczba gospodarstw	115	954	360	98
	średni błąd predykcji dla modelu z trendem	111,996	123,965	114,634	142,324
	średni błąd predykcji dla modelu bez trendu	100,328	126,747	100,845	110,422
Rzepak i rzepik	średnie $\hat{\beta}_i$	-0,716	-1,514	-0,684	-0,577
	odchylenie standardowe $\hat{\beta}_i$	1,880	1,919	1,787	1,530
	liczba gospodarstw	389	1109	318	155
	średni błąd predykcji dla modelu z trendem	12,922	12,549	9,816	10,967
	średni błąd predykcji dla modelu bez trendu	8,910	11,277	8,362	8,213

Źródło: obliczenia własne na podstawie danych FADN.

WNIOSKI

W przypadku przeciętnych plonów zbóż na poziomie gospodarstwa nie ma wyraźnego trendu. Natomiast w przypadku kukurydzy na ziarno we wszystkich regionach Polski zaobserwowano poprawę plonowania. Również buraki cukrowe wykazały poprawę plonowania. Jedyne w przypadku rzepaku odnotowano w ostatnich latach pogorszenie przeciętnego plonowania.

Uzyskane wyniki podważają zasadność stosowania funkcji trendu do prognozowania przyszłych plonów na poziomie gospodarstwa rolnego. Przeciętnie, biorąc pod uwagę wszystkie rozpatrywane rośliny uprawne, zastosowanie funkcji trendu do predykcji plonów powoduje zwiększenie średniego błędu predykcji o około 27% w porównaniu z użyciem średniej arytmetycznej jako prognozy przyszłych plonów. Nawet w przypadku upraw, takich jak: kukurydza na ziarno, buraki cukrowe i rzepak, dla których zaobserwowano przeciętne poziomy współczynnika kierunkowego wyraźnie różne od zera, użycie funkcji trendu do predykcji prowadzi do zwiększenia średniego błędu predykcji o około 12% dla kukurydzy i buraków cukrowych oraz 26% dla rzepaku.

LITERATURA

- Goraj Lech, Malanowska Beata, Osuch Dariusz, Sierański Witold, 2012: *Opis realizacji planu wyboru próby gospodarstw rolnych dla Polskiego FADN w 2012 roku*, Polski FADN.
- Kobus Paweł, Wicki Ludwik, 2010: *Variability of Cereals Yields in Poland in the Period of 1968-2008*, „Economic Science for Rural Development”, nr 21.
- Marra Michele C., Schurle Bryan W. 1994: *Kansas wheat yield risk measures and aggregation: a meta-analysis approach*, „Journal of Agricultural and Resource Economics”, 19(1).
- Moss Charles B. 2010: *Risk, uncertainty and the agricultural firm*, World Scientific Publishing Co. Pte. Ltd.

Peltonen-Sainio Pirjo, Jauhiainen Lauri, Laurila Ilkka P. 2009: *Cereal yield trends in northern European conditions: Changes in yield potential and its realisation*, „Field Crops Research”, vol. 110, Issue 1, 5 January 2009, s. 85-90.

Rocznik statystyczny rolnictwa 2013. 2013: GUS, Zakład Wydawnictw Statystycznych, Warszawa.

Wicki Ludwik 2010: *Efekty upowszechniania postępu biologicznego w produkcji roślinnej*, Wydawnictwo SGGW, Warszawa.

Paweł Kobus

THE PROBLEM OF TRENDS ELIMINATION IN THE CROP PRODUCTION RISK ASSESSMENT

Summary

Historical data of the crops yield confirms the existence of a systematic change of the average crops yield at the country level. During assessment of the production risk the overall variability of yield production is usually reduced by the variability attributed to a trend. The aim of the study was to assess the validity of such procedure. To do this, yield trends of basic crops in Poland were estimated. Afterward, the quality of the yield prediction based on the trends and the quality of the prediction based on the arithmetic mean were compared. As a measure of the prediction quality the square root of the mean squared prediction error was adopted. The study has shown that the use of the trend function leads, on average, to the lowering prediction quality by about 27%.

Adres do korespondencji:

Dr Paweł Kobus

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

Wydział Nauk Ekonomicznych

Katedra Ekonomiki Rolnictwa i Międzynarodowych Stosunków Gospodarczych

ul. Nowoursynowska 166, 02-787 Warszawa

e-mail: pawel_kobus@sggw.pl