

Modelowanie Bezrobocia w Czechach

Projekt zaliczeniowy z przedmiotu Ekonometria I

Hanna Malash 122514

1. PRZYGOTOWANIE TEORETYCZNEJ KONCEPCJI MODELU

CEL PROJEKTU

Celem projektu jest opracowanie modelu ekonometrycznego dla zbadania i opisanie zmian procentu bezrobocia w Czechach oraz wykorzystanie go do przeprowadzenia prognozy na dwa lata, na podstawie danych zebranych z oficjalnych źródeł, takich jak Eurostat.

DYSKUSJA NA TEMAT DOBORU ZMIENNYCH OBJAŚNIAJĄCYCH

Zmienną objaśnianą w modelu jest stopa bezrobocia w Czechach. W procesie modelowania uwzględniono następujące potencjalne czynniki wpływające na zmienność tej zmiennej:

- **Podstawowe czynniki ekonomiczne:**
 - **Produkt Krajowy Brutto (GDP):** Wartość wyrażona w milionach euro w cenach bieżących (dane kwartalne z Eurostatu). Wzrost PKB może wpływać na spadek stopy bezrobocia poprzez tworzenie nowych miejsc pracy.
 - **Produkcja przemysłowa (Production in Industry):** Wskaźnik (index, rok bazowy = 2015). Zmiany w aktywności sektora przemysłowego wskazują na jego wpływ na rynek pracy.
 - **Produkcja w budownictwie (Production in Construction):** Wskaźnik (index, rok bazowy = 2015). Zmiany w aktywności sektora budowlanego odzwierciedlają jego wpływ na zatrudnienie, szczególnie w obszarach związanych z infrastrukturą.
 - **Wskaźnik kosztów pracy (Labour Cost Index):** Index (rok bazowy = 2015). Mierzy zmiany w kosztach pracy, które mogą wpływać na decyzje pracodawców o zatrudnieniu.
- **Czynniki demograficzne i społeczne:**
 - **Populacja w wieku produkcyjnym (Population):** Liczba ludności zdolnej do pracy (w tysiącach osób). Wskazuje na podaż pracy na rynku.
 - **Nowi pracownicy (Job Starters):** Liczba osób rozpoczynających pracę (index, rok bazowy = 2015). Reprezentuje dynamikę wejścia nowych osób na rynek pracy.
- **Wymiar międzynarodowy:**
 - **Kurs walutowy (EUR/CZK):** Średni kwartalny kurs wymiany waluty. Wpływa na eksport, a co za tym idzie, na zatrudnienie w sektorze eksportowym.

ODRZUCONE ZMIENNE

W procesie modelowania zmuszeni byliśmy odrzucić:

- **Job Vacancy:**
 - Typ: *Index (2015 = 100)*
 - Pomimo wysokiej korelacji z bezrobociem (-0.916), została odrzucona ze względu na redundancję z innymi zmiennymi.

- **Kurs walutowy (EUR/CZK):**

- Typ: *Średni kwartalny kurs wymiany waluty*
- Korelacja z bezrobociem była niska (0.121), a jej wpływ na rynek pracy marginalny.

- **Godziny pracy (Hours Worked):**

- Typ: *Średnia liczba godzin pracy tygodniowo*
- Słaba korelacja z bezrobociem (0.092) wskazywała na brak istotności tej zmiennej.

- **Nowi pracownicy (Job Starters):**

- Typ: *Index (2015 = 100)*
- Ostatecznie odrzucona ze względu na problemy z redundancją wobec innych zmiennych.

OSTATECZNE ZMIENNE W MODELU

Po przeprowadzeniu wstępnych analiz, model uwzględnił następujące zmienne:

1. **Produkcja przemysłowa (Production in Industry):**

- Typ: *Index (2015 = 100)*
- Uwzględniona jako kluczowy wskaźnik aktywności gospodarczej z wysoką negatywną korelacją z bezrobociem (-0.867).

2. **Produkt Krajowy Brutto (GDP):**

- Typ: *Miliony euro (w cenach bieżących)*
- Uwzględniony jako istotny wskaźnik ogólnej kondycji gospodarki, z umiarkowaną korelacją z bezrobociem (-0.789).

3. **Wskaźnik kosztów pracy (Labour Cost Index):**

- Typ: *Index (2015 = 100)*
- Wskazuje na zmiany kosztów zatrudnienia, które mogą wpływać na decyzje o zatrudnieniu, korelacją z bezrobociem (-0.633).

4. **Populacja (Population):**

- Typ: *Liczba w tysiącach osób*
- Uwzględniona jako zmienna demograficzna opisująca podaż pracy, korelacją z bezrobociem (-0.889).

TRANSFORMACJE ZMIENNYCH:

Podczas budowy modelu dla zmian stopy bezrobocia w Czechach zastosowaliśmy transformacje zmiennych w celu poprawy ich właściwości statystycznych oraz dopasowania modelu do danych. Kluczowe działania obejmowały:

- **Unemployment:** Zmienna była niestacjonarna, zastosowano pierwsze różnice ($d_Unemployment$), aby usunąć niestacjonarność i poprawić właściwości reszt.
- **GDP:** Zmienna niestacjonarna, zastosowano pierwsze różnice (d_GDP), aby uzyskać stacjonarność i poprawić dopasowanie modelu.
- **Labour Cost Index:** Ujęto z opóźnieniem o jeden kwartał ($LabourCostIndex_1$), aby uwzględnić opóźniony wpływ kosztów pracy na bezrobocie.
- **Production in Construction:** Zastąpiono zmienną **Production in Industry**, ponieważ lepiej wyjaśniała zmienność stopy bezrobocia i wykazywała silniejszą korelację z nią (-0.867).
- **Population:** Zmienna niestacjonarna, zastosowano pierwsze różnice ($d_Population$), aby zachować stacjonarność i umożliwić jej wykorzystanie w modelu.

Model wyjściowy:

gretl: model 6

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Model 6: Estymacja KMNK, wykorzystane obserwacje 2009:2-2021:4 (N = 51)
Zmienna zależna (Y): Unemployment

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	10,6353	0,702454	15,14	1,72e-19 ***
LabourCostIndex	-0,0465604	0,0235346	-1,978	0,0539 *
JobVacancy	-0,727324	0,0564940	-12,87	7,39e-17 ***
ProductioninCons-	-0,00767208	0,00828256	-0,9263	0,3591
GDP	-5,85612e-05	1,23055e-05	-4,759	1,96e-05 ***

Sredn.aryt.zm.zależnej 4,672549 Odch.stand.zm.zależnej 2,010381
Suma kwadratów reszt 6,091700 Błąd standardowy reszt 0,363907
Wsp. determ. R-kwadrat 0,969855 Skorygowany R-kwadrat 0,967234
F(4, 46) 369,9926 Wartość p dla testu F 2,45e-34
Logarytm wiarygodności -18,18095 Kryt. inform. Akaike'a 46,36191
Kryt. bayes. Schwarza 56,02103 Kryt. Hannana-Quinna 50,05295
Autokorel.reszt - rho1 0,532636 Stat. Durbin-Watsona 0,909856

Wyłączając stałą, największa wartość p jest dla zmiennej 6 (ProductioninConstruction)

Model ostateczny:

gretl: model 7

Plik Edycja Testy Zapisz Wykresy Analiza LaTeX

Model 7: Estymacja KMNK, wykorzystane obserwacje 2009:2-2021:4 (N = 51)
Zmienna zależna (Y): d_Unemployment

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	1,41729	0,314323	4,509	4,47e-05 ***
ProductioninIndu-	-0,0166845	0,00338105	-4,935	1,09e-05 ***
d_GDP	4,61266e-05	1,31905e-05	3,497	0,0011 ***
LabourCostIndex_1	0,0434427	0,0115367	3,766	0,0005 ***
d_Population	-0,00521708	0,00115901	-4,501	4,58e-05 ***

Sredn.aryt.zm.zależnej -0,062745 Odch.stand.zm.zależnej 0,308520
Suma kwadratów reszt 2,251312 Błąd standardowy reszt 0,221227
Wsp. determ. R-kwadrat 0,526957 Skorygowany R-kwadrat 0,485823
F(4, 46) 32,81071 Wartość p dla testu F 4,37e-07
Logarytm wiarygodności 7,202104 Kryt. inform. Akaike'a -4,404208
Kryt. bayes. Schwarza 5,254921 Kryt. Hannana-Quinna -0,713169
Autokorel.reszt - rho1 0,265446 Stat. Durbin-Watsona 1,408475

POWODY ZMIAN:

- Usunięcie niestacjonarności (Unemployment, GDP, Population) w celu uzyskania wiarygodnych wyników.
- Ujęcie opóźnienia w Labour Cost Index poprawiło dopasowanie modelu i uwzględniło opóźnione efekty.
- Zastąpienie Production in Construction zmienną Production in Industry poprawiło zdolność modelu do wyjaśniania zmienności.

SKUTKI ZMIAN:

Transformacje zmiennych, takie jak różnicowanie niestacjonarnych zmiennych (**Unemployment**, **GDP**, **Population**) oraz wprowadzenie opóźnienia w **Labour Cost Index**, pozwoliły na poprawę stacjonarności i eliminację problemów z autokorelacją reszt (statystyka Durbin-Watsona wzrosła do 1.408). W modelu pierwotnym zbyt wysoki (R^2 : 0.969) sugerował przeuczony model, co zostało skorygowane w modelu finalnym (R^2 : 0.526). Zamiana **Production in Construction** na **Production in Industry** poprawiła istotność statystyczną i zdolność wyjaśniania zmienności, zgodnie z rolą sektora przemysłowego w kształtowaniu rynku pracy w Czechach

STATYSTYKI OPISOWE ZMIENNYCH:



Zmienna	Średnia	Mediana	Odchylenie standardowe (S.D.)	Minimalna	Maksymalna
Production in Industry	101.4	101.8	11.91	87.9	118.4
Labour Cost Index	4.315	4	3.231	-2.8	10.1
d_Population	4.671	5	35.5	-96.7	82.4
d_GDP	635.2	1269	3122	-6539	7372

Skośność: Dla wszystkich zmiennych mamy do czynienia ze skośnością lewostronną. Jej wartość waha się między -0,01 a -0,53

Kurtoza: Dla wszystkich zmiennych mamy do czynienia z rozkładami spłaszczonymi (leptokurtycznymi). Jej wartość waha się między -1,14 a 0,54.

Normalność rozkładu: Po wykonaniu testu Jarque'a-Bera dla reszt modelu (uhat1) okazuje się, że wartość p wynosi 0,040061 i jest mniejsza od poziomu 0,05. W związku z tym mamy podstawy do odrzucenia hipotezy zerowej o normalnym rozkładzie składnika losowego. Dodatkowe testy Shapiro-Wilka, Doornika-Hansena oraz Lillieforsa również wskazują na podobne wyniki, gdzie większość wartości p jest zbliżona do poziomu krytycznego, ale wciąż nie pozwala jednoznacznie zaakceptować normalności rozkładu.

Test na normalność rozkładu uhat1:

```
Test Doornika-Hansena = 5,00207, z wartością p 0,0820002
Test Shapiro-Wilka = 0,955229, z wartością p 0,0522862
Test Lillieforsa = 0,0947585, z wartością p ~ 0,29
Test Jarque'a-Bera = 6,43471, z wartością p 0,0400609
```

Stacjonarność: P- value w teście Dickeya-Fullera dla żadnej ze zmiennych objaśniających nie przekracza 0,05. Należy stwierdzić, że nie ma podstaw do odrzucenia hipotezy zerowej i uznać, że wszystkie zmienne są stacjonarne.

2. WERYFIKACJA STATYSTYCZNA MODELU

DOBÓR SPECYFIKACJI MODELU

Model jest statyczny, ponieważ zmienna zależna $\Delta Unemployment_t$ jest wyjaśniana wyłącznie przez bieżące wartości zmiennych objaśniających oraz jedną zmienną z opóźnieniem ($LabourCostIndex_{t-1}$). Brak opóźnionych wartości zmiennej zależnej wskazuje na brak dynamicznych efektów, charakterystycznych dla modeli ADL czy ECM.

INTERPRETACJA PARAMETRÓW

$$\Delta Unemployment_t = \beta_0 + \beta_1 \cdot Production_in_Industry_t + \beta_2 \cdot \Delta GDP_t + \beta_3 \cdot LabourCostIndex_{t-1} + \beta_4 \cdot \Delta Population_t + \epsilon_t$$

- Wzrost indeksu **produkcji przemysłowej** o 1 punkt powoduje spadek zmian bezrobocia o 0,01668 p.p., ceteris paribus.
- Wzrost **liczby ludności** o 1000 osób powoduje spadek zmian bezrobocia o 0,005 p.p., ceteris paribus
- Wzrost **wskaźnika kosztów pracy** o 1 punkt w poprzednim okresie powoduje wzrost zmian bezrobocia o 0,043 punktu procentowego, ceteris paribus.
- Wzrost **pierwszych różnic PKB** o 1 milion euro powoduje wzrost zmian bezrobocia o 0,000047 punktu procentowego, ceteris paribus.

WERYFIKACJA WŁASNOŚCI MODELU

a) **Istotność zmiennych** – Indeks produkcji przemysłowej, pierwsze różnice populacji, opóźniony wskaźnik kosztów pracy oraz pierwsze różnice PKB mają p-value znacznie mniejsze niż 0,05, więc stwierdzamy, że te zmienne są istotne statystycznie. Ich włączenie do modelu jest uzasadnione zarówno statystycznie, jak i ekonomicznie, co potwierdza ich wpływ na dynamikę bezrobocia w Czechach.

b) **Dopasowanie modelu do danych** – Współczynnik determinacji R^2 wynosi 0,52, co oznacza, że model wyjaśnia zmiany bezrobocia w Czechach w 52%. Wartość ta wskazuje na umiarkowaną zdolność modelu do odzwierciedlenia dynamiki danych.

c) **Poprawność specyfikacji** – p-value otrzymane w teście RESET wynosi 1,00, co nie daje nam podstaw do odrzucenia hipotezy zerowej. Stwierdzamy, że model ma poprawną specyfikację.

d) **Współliniowość** – Wartość żadnego z czynników wariancji inflacji nie zbliża się do 10, więc możemy stwierdzić, że problem współliniowości nie występuje.

e) **Rozkład reszt modelu** – p-value w teście normalności rozkładu reszt wynosi 0,040061, więc odrzucamy hipotezę zerową i uznajemy hipotezę alternatywną mówiącą o braku rozkładu normalnego.

f) **Heteroskedastyczność i autokorelacja** – Wyniki testu Breuscha-Godfreya wskazują, że p-value wynoszące 0,398 dla LMF, 0,331 dla TR^2 oraz 0,26 dla testu Ljunga-Boxa są większe niż 0,05, co oznacza brak podstaw do odrzucenia hipotezy zerowej o braku autokorelacji reszt modelu. Test White'a dla heteroskedastyczności wskazuje na p-value wynoszące 0,352, co również nie daje podstaw do odrzucenia hipotezy zerowej o jednorodności wariancji reszt.

3. PRZEPROWADZENIE ANALIZY WŁASNOŚCI PROGNOSTYCZNYCH MODELU

MIARY BŁĘDÓW PROGNOZY EX-POST

Miary dokładności prognoz ex post wykorzystano 7 obserwacji

Średni błąd predykcji	ME =	0,10664
Pierwiastek błędu średniokwadr.	RMSE =	0,25423
Średni błąd absolutny	MAE =	0,22936
Współczynnik Theila (w procentach)	U2 =	1,5847
Udział obciążoności predykcji	UM =	0,17596
Udział niedost. elastyczności	UR =	0,67943
Udział niezgodności kierunku	UD =	0,14461

$ME = 0,11$ – biorąc pod uwagę zróżnicowanie wartości bezrobocia, jest to niewielka wartość wskazująca na niewielkie obciążenie prognozy i nieznaczne zawyżanie wyników.

$RMSE = 0,25$ – prognoza myli się średnio o 0,25 p.p., co można uznać za zadowalający wynik, biorąc pod uwagę charakter danych.

$MAE = 0,23$ – średni błąd bezwzględny wynosi 0,23, co potwierdza, że prognoza jest stosunkowo precyzyjna i dobrze dopasowana do rzeczywistych wartości.

$UM = 0,17596$ - znacząca część błędu prognozy wynika z systematycznego obciążenia modelu. Wskazuje to na pewne odchylenie w przewidywanych wartościach w stosunku do rzeczywistości.

$UR = 0,67943$ - większość błędu prognozy wynika z niedostatecznego dostosowania modelu do zmienności danych. Model może nie być wystarczająco elastyczny, aby dokładnie odzwierciedlić dynamikę badanych zjawisk.

$UD = 0,14461$ - wskazuje, że około 14,5% błędów wynika z błędnych przewidywań kierunku zmian. Jest to stosunkowo niski udział, co oznacza, że model w większości przypadków poprawnie przewiduje kierunek zmian.

Prognoza obejmuje okres od pierwszego kwartału 2022 roku do trzeciego kwartału 2023 roku i charakteryzuje się bardzo dobrymi miarami błędów ex-post, co wskazuje na wysoką jakość dopasowania modelu, choć pewne aspekty, takie jak elastyczność modelu, mogą wymagać dalszej poprawy.

MIARY BŁĘDÓW PROGNOZY EX-POST



Prognoza jest w miarę dobrze dopasowana do rzeczywistych danych, szczególnie w pierwszej części próby. W późniejszym okresie widać pewne odchylenia, zwłaszcza w momentach większej zmienności, jednak prognoza pozostaje w granicach wyznaczonego przedziału ufności. Świadczy to o wysokiej jakości prognozy, przy czym większa niepewność w ostatnich okresach jest typowa dla prognoz dynamicznych.

Wyniki miar prognoz ex-post, takich jak $RMSE = 0,25$ i $MAE = 0,23$, potwierdzają wysoką precyzję prognoz. $ME = 0,11$ wskazuje na niewielkie obciążenie prognozy, a współczynnik $U2$ Theila równy $1,5847$ świadczy o akceptowalnym dopasowaniu modelu. Wartości $UM = 0,17596$, $UR = 0,67943$ oraz $UD = 0,14461$ wskazują, że większość błędów prognozy wynika z niedostatecznego dopasowania modelu do zmienności danych, choć odchylenia w przewidywaniu kierunku zmian są stosunkowo niewielkie.

Podsumowując, model jest jednym z możliwych narzędzi do analizy i prognozowania zmian bezrobocia w Czechach. Dostarcza wiarygodnych wyników, cechuje się dobrą specyfikacją oraz precyzją przewidywań, co czyni go odpowiednim narzędziem do badania dynamiki rynku pracy.

4. PODSUMOWANIE

Model ekonometryczny dla zmian stopy bezrobocia w Czechach skutecznie wyjaśnia dynamikę tej zmiennej, co potwierdzają dobre wyniki testów diagnostycznych i prognoz ex-post. Specyfikacja modelu, oparta na pierwszych różnicach zmiennych, uwzględnia takie determinanty, jak produkcja przemysłowa (Production in Industry), zmiany produktu krajowego brutto (GDP), wskaźnik kosztów pracy z opóźnieniem (Labour Cost Index) oraz zmiany populacji (Population). Współczynnik determinacji modelu ($R^2 = 0,526957$) wskazuje, że model wyjaśnia około 53% zmienności zmian bezrobocia.

Test RESET wykazał, że p-value wynoszące 0,74 nie daje podstaw do odrzucenia hipotezy zerowej o poprawnej specyfikacji modelu. Test White'a (p-value = 0,352) potwierdził brak heteroskedastyczności w modelu. Analiza współliniowości wskazała na niewielkie wartości współczynników determinacji regresji pomocniczych, co świadczy o braku istotnych problemów współliniowości.