

Komentarz wstępu:

W poprzedniej pracy domowej okazało się, że preferowanym modelem był model SEM. Pozostałe modele z jednym źródłem procesu przestrzennego zostały odrzucone – w modelu SAR Rho było statystycznie równe 0, a w modelu SLX reszty wykazywały autokorelację przestrzenną oraz występowały dziwne schematy wymienności pomiędzy zmiennymi a opóźnieniami tych zmiennych.

Macierz W w każdym modelu to macierz oparta na odwróconych kwadratach odległości.

Model SARAR:

```
Type: sac
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -18.71118   176.23299 -0.1062  0.915445
spatial_data3$gdp_per_capita    3.47079    8.35856  0.4152  0.677968
spatial_data3$pop_density      0.28395    0.04837  5.8703 4.349e-09
spatial_data3$n_of_police_per_100k 53.93202   20.24137  2.6644 0.007712
spatial_data3$emp_share    199.08807   428.26823  0.4649  0.642026

Rho: -0.17563
Asymptotic standard error: 0.1244
z-value: -1.4118, p-value: 0.15801
Lambda: 0.53213
Asymptotic standard error: 0.14031
z-value: 3.7926, p-value: 0.0001491

LR test value: 8.111, p-value: 0.017327

Log likelihood: -427.0528 for sac model
ML residual variance (sigma squared): 18495, (sigma: 136)
Number of observations: 67
Number of parameters estimated: 8
AIC: 870.11, (AIC for lm: 874.22)
```

Rho jest statystycznie równe 0 przy każdym sensownym poziomie istotności. Kryterium AIC jest wyższe niż w przypadku modelu SEM, a więc model SARAR nie dodaje żadnych cennych własności, które pomogłyby przy ocenie danego zjawiska.

Lambda jest statystycznie różna od 0, podobnie jak w modelu SEM.

Podobnie jak w modelu SEM, zmienną istotną jest gęstość populacji (pop_density) oraz liczba posterunków policji na 100tys mieszkańców (n_of_police_per_100k). Oszacowania przy tych parametrach są bardzo zbliżone do oszacowań z modelu SEM.

LR test SARAR vs SEM (SEM jest preferowany w modelach z jednym źródłem procesu przestrzennego)

P-value dla testu: 0,44

H0 broni się, co oznacza, że możemy odpuścić model SARAR na rzecz modelu SEM.

Reszty w modelu

Moran I statistic standard deviate = -0.4897, p-value = 0.6878
--

Test Morana wskazuje, że nie występuje autokorelacja przestrzenna reszt w modelu.

Wnioski

Model z jednym źródłem procesu przestrzennego SEM jest preferowany względem modelu SARAR.

Model SDM

```
Type: mixed
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    9.3802e+02  3.3955e+02  2.7625  0.005735
spatial_data3$gdp_per_capita  5.7202e-01  8.1436e+00  0.0702  0.944002
spatial_data3$pop_density    3.0913e-01  4.3776e-02  7.0617  1.645e-12
spatial_data3$n_of_police_per_100k  3.5004e+01  2.0924e+01  1.6729  0.094353
spatial_data3$emp_share    1.7528e+02  4.1498e+02  0.4224  0.672754
lag.spatial_data3$gdp_per_capita  1.6765e+01  1.4042e+01  1.1939  0.232506
lag.spatial_data3$pop_density   -2.9091e-02  9.1073e-02  -0.3194  0.749404
lag.spatial_data3$n_of_police_per_100k -1.3971e+02  4.3502e+01  -3.2115  0.001320
lag.spatial_data3$emp_share   -2.0095e+03  8.6905e+02  -2.3123  0.020764

Rho: 0.25702, LR test value: 3.2356, p-value: 0.072053
Asymptotic standard error: 0.14802
z-value: 1.7364, p-value: 0.082484
Wald statistic: 3.0153, p-value: 0.082484

Log likelihood: -422.4328 for mixed model
ML residual variance (sigma squared): 17256, (sigma: 131.36)
Number of observations: 67
Number of parameters estimated: 11
AIC: 866.87, (AIC for lm: 868.1)
LM test for residual autocorrelation
test value: 0.067017, p-value: 0.79573
```

Współczynnik Rho w modelu SDM jest wciąż statystycznie równy 0, jednakże przy poziomie istotności na poziomie 0,05.

Podobnie jak w przypadku modelu SLX, istotne są oszacowania parametrów przy zmiennych pop_density, przy opóźnieniu przestrzennym liczby posterunków policji na 100tys mieszkańców oraz opóźnieniu przestrzennym zmiennej dot. udziału osób pracujących. Kierunek zależności wydaje się uzasadniony, jednakże oszacowania przy wyżej wymienionych zmiennych nieopóźnionych są nieistotne. Jest to w tym przypadku mało akceptowalne i raczej niepoprawne. Możemy powiedzieć że model jest przeuczony, w modelu uwzględnione są zmienne które mogą wywoływać współliniowość, a co za tym idzie zaciemniają precyzję szacunków poszczególnych parametrów.

Kryterium AIC jest niższe w tym modelu w porównaniu z modelem SARAR, jednak wiemy, że kryterium AIC ma tendencję do wybierania modelu o zbyt dużej liczbie parametrów.

LR test SDM vs SEM (SEM jest preferowany w modelach z jednym źródłem procesu przestrzennego)

P-value dla testu: 0.004

Odrzucamy H0 na rzecz H1 przy każdym sensownym poziomie istotności. Oznacza to, że model SDM jest preferowany względem modelu SEM.

Reszty w modelu

Moran I statistic standard deviate = 0.29344, p-value = 0.3846
--

Test Morana wskazuje, że nie występuje autokorelacja przestrzenna reszt w modelu.

Wnioski

Biorąc pod uwagę wartość AIC oraz wnioski z testu LR można by było stwierdzić, że model SDM jest modelem preferowanym względem modelu SEM, który okazał się najlepszy w modelowaniu danego zjawiska w klasie modeli z jednym źródłem procesu przestrzennego. Jednakże biorąc pod uwagę wnioski na temat interpretacji oraz braku istotności kluczowych parametrów w modelu SDM stwierdzamy, że model ten jest przeuczony i modelem preferowanym jest model SEM.

Model SDEM

```
Type: error
Coefficients: (asymptotic standard errors)

              Estimate Std. Error z value Pr(>|z|)
(Intercept)      1.0909e+03  3.9490e+02  2.7625  0.005737
spatial_data3$gdp_per_capita  1.0016e+00  8.1706e+00  0.1226  0.902438
spatial_data3$pop_density    3.0909e-01  4.5788e-02  6.7505  1.473e-11
spatial_data3$n_of_police_per_100k  2.6343e+01  2.0792e+01  1.2670  0.205172
spatial_data3$emp_share      7.1041e+01  4.0986e+02  0.1733  0.862392
lag.spatial_data3$gdp_per_capita  1.9794e+01  1.4773e+01  1.3399  0.180291
lag.spatial_data3$pop_density    6.3166e-02  8.3609e-02  0.7555  0.449954
lag.spatial_data3$n_of_police_per_100k -1.3501e+02  4.6145e+01 -2.9257  0.003437
lag.spatial_data3$emp_share     -2.2484e+03  9.2654e+02 -2.4267  0.015236

Lambda: 0.30552, LR test value: 3.8084, p-value: 0.050997
Asymptotic standard error: 0.15268
z-value: 2.0011, p-value: 0.045384
Wald statistic: 4.0043, p-value: 0.045384

Log likelihood: -422.1464 for error model
ML residual variance (sigma squared): 16990, (sigma: 130.35)
Number of observations: 67
Number of parameters estimated: 11
AIC: 866.29, (AIC for lm: 868.1)
```

Parametr Lambda jest statystycznie różny od zera. Po raz kolejny występują dziwne schematy wymienności pomiędzy zmiennymi a opóźnieniami tych zmiennych (w przypadku tych samych zmiennych co w modelu SDM oraz SLX).

Kryterium AIC jest niemal identyczne jak w modelu SDM.

LR test SDEM vs SEM (SEM jest preferowany w modelach z jednym źródłem procesu przestrzennego)

P-value dla testu: 0.0033

Odrzucamy H_0 na rzecz H_1 przy każdym sensownym poziomie istotności. Oznacza to, że model SDM jest preferowany względem modelu SEM.

Reszty w modelu

Moran I statistic standard deviate = -0.012869, p-value = 0.5051
--

Test Morana wskazuje, że nie występuje autokorelacja przestrzenna reszt w modelu.

Wnioski

Odrzucamy model SDEM na rzecz modelu SEM, z tych samych powodów co w przypadku modelu SDM.

Kryterium AIC					
SAR	SEM	SLX	SARAR	SDM	SDEM
875,36	869,73	868,1	870,11	866,87	866,29

Który z sześciu modeli jest preferowany?

Na podstawie wcześniejszej analizy preferowanym modelem z jednym źródłem procesu przestrzennego jest model SEM.

We wszystkich modelach z dwoma źródłami nie wystąpiła autokorelacja przestrzenna reszt, podobnie jak w przypadku modelu SEM.

Jeśli chodzi o porównanie modelu SARAR z SEM widzimy, że wartość kryterium AIC jest niższa dla modelu SEM, w modelu SARAR parametr Rho jest statystycznie równy 0 oraz LR test SARAR vs SEM wykazał, że model SEM jest modelem preferowanym.

W przypadku porównania modeli SDM i SEM – kryterium AIC jest niższe dla modelu SDM, LR test SDM vs SEM wykazał, że model SDM jest modelem preferowanym. Jednakże parametr Rho w modelu SDM jest statystycznie równy 0 przy poziomie istotności 0,05, ponadto okazuje się, że model SDM jest przeuczony – parametry przy oszacowaniach zmiennych są nieistotne, natomiast parametry przy niektórych zmiennych opóźnionych są istotne. Jest to w tym przypadku nieakceptowalne i dlatego w tym porównaniu lepiej wypada model SEM.

Analogiczna sytuacja występuje przy porównaniu modeli SDEM i SEM. Przez to że model SDEM wykazuje dziwne schematy wymienności pomiędzy zmiennymi a opóźnieniami tych zmiennych odrzucamy go na rzecz modelu SEM mimo faktu, że kryterium AIC jest niższe dla modelu SDEM oraz LR test SDEM vs SEM wykazał, że model SDEM jest preferowany.

Biorąc pod uwagę wszystkie porównania, a także wyniki z poprzedniej pracy domowej, model SEM jest modelem preferowanym.