

ESTYMACJA WSKAŹNIKA BARDZO NISKIEJ INTENSYWNOŚCI PRACY Z WYKORZYSTANIEM MODELU FAYA-HERRIOTA I JEGO ROZSZERZEŃ

Łukasz Wawrowski, Maciej Beręsewicz

12.06.2015

Urząd Statystyczny w Poznaniu, Uniwersytet Ekonomiczny w Poznaniu

PLAN PREZENTACJI

Wstęp

Zastosowane modele

Wyniki

Podsumowanie

WSTĘP

Monitorowanie postępów strategii Europa 2020, mającej na celu działania na rzecz włączenia społecznego, wymaga dokładnych i szczegółowych danych statystycznych.

Rozpoznanie możliwości dezagregacji wskaźnika bardzo niskiej intensywności pracy w gospodarstwach domowych na poziom województw

Definicja

Odsetek osób w wieku 0-59 lat mieszkających w gospodarstwach domowych o bardzo niskiej intensywności pracy, tzn. takich, w których osoby dorosłe (w wieku 18–59 lat) w minionym roku przepracowały mniej niż 20% ich całkowitego potencjału pracy

Wskaźnik odnosi się do liczby miesięcy, w czasie których wszyscy członkowie gospodarstwa domowego w wieku aktywności zawodowej, pracowali w roku odniesienia dochodu (Y-1), jako proporcja całkowitej liczby miesięcy, która teoretycznie mogła zostać przepracowana, do liczby miesięcy faktycznie przepracowanych. Jako osoby w wieku aktywności zawodowej uznaje się osoby w wieku 18–59 lat, które nie są dziećmi na utrzymaniu.

ZASTOSOWANE MODELE

Podstawowe założenia

1. Estymacja oparta o schemat losowania
2. Nie zakłada istnienia efektów losowych dla domen
3. Nie zakłada skorelowania zmiennych y_k, y_l
4. Wariancja rośnie wraz ze spadkiem wielkości próby n_d
5. Brak oszacowań dla domen niereprezentowanych w próbie $n_d = 0$

Estymator bezpośredni dany jest następującym wzorem:

$$\hat{y}_{dt}^{dir} = \frac{\sum_{i=1}^{n_{dt}} y_{idt} w_{idt}}{\sum_{i=1}^{n_{dt}} w_{idt}} \quad (1)$$

gdzie $d = 1, 2, \dots, D$ i $t = 1, 2, \dots, T$.

Podstawowe założenia

1. Dynamiczny model zakładający liniowy model Fay'a-Herriot'a dla domen $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, który w tym przypadku ma postać $y_{dt} = \mathbf{X}\boldsymbol{\beta} + u_{1d} + u_{2dt}$.
2. Stacjonarność szeregu czasowego.
3. Zakłada istnienie dwóch efektów losowych - dla domeny u_{1d} ($u_{1d} \sim N(0, \sigma_1^2)$), gdzie u_{1d} jest efektem losowym w czasie $t = 1$ oraz efektu czasowego dla domeny $u_{2dt} = \rho_2 u_{2d,t-1} + \epsilon_{2dt}$ ($\epsilon_{2dt} \sim N(0, \sigma_2^2)$).

EBLUP dla tego modelu ma postać

$$\hat{y}_{dt}^{\text{EBLUP RY}} = \mathbf{x}'_{dt} \hat{\boldsymbol{\beta}} + (\hat{\sigma}_1^2 \mathbf{1}'_T + \hat{\sigma}_2^2 \boldsymbol{\gamma}'_t) (\boldsymbol{\Sigma}_d + \hat{\sigma}_1^2 \boldsymbol{\Gamma} + \hat{\sigma}_2^2 \mathbf{J}_T)^{-1} (\mathbf{y}_d - \mathbf{X}_d \boldsymbol{\beta}) \quad (2)$$

gdzie $d = 1, 2, \dots, D$ i $t = 1, 2, \dots, T$.

Podstawowe założenia

1. Dynamiczny model zakładający liniowy model Fay'a-Herriot'a dla domen $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, który w tym przypadku ma postać $\mathbf{y}_{dt} = \mathbf{X}\boldsymbol{\beta} + \rho^{t-1}\mathbf{u}_{1d} + \mathbf{u}_{2dt}$.
2. Zakłada niestacjonarność szeregu czasowego.
3. Zakłada istnienie dwóch efektów losowych skorelowanych w czasie - dla domeny $\rho^{t-1}\mathbf{u}_{1d}$, gdzie \mathbf{u}_{1d} jest efektem losowym w czasie $t = 1$ ($u_{1d} \sim N(0, \sigma_1^2)$) oraz efektu czasowego dla domeny $\mathbf{u}_{2dt} = \rho_2\mathbf{u}_{2d,t-1} + \boldsymbol{\epsilon}_{2dt}$ ($\boldsymbol{\epsilon}_{2dt} \sim N(0, \sigma_2^2)$).

EBLUP dla tego modelu ma postać

$$\hat{\mathbf{y}}_{dt}^{\text{EBLUP FPD}} = \mathbf{x}'_{dt}\hat{\boldsymbol{\beta}} + (\hat{\sigma}_1^2\mathbf{1}'_T + \hat{\sigma}_2^2\boldsymbol{\gamma}'_t)'V^{-1}(\mathbf{y}_d - \mathbf{X}_d\boldsymbol{\beta}) \quad (3)$$

gdzie $d = 1, 2, \dots, D$ i $t = 1, 2, \dots, T$.

Podstawowe założenia

1. Zakłada liniowy model Fay'a-Herriot'a dla domen $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$, który w tym przypadku ma postać $\mathbf{y}_{dt} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_{1d} + \mathbf{u}_{2dt}$.
2. Autoregresję AR(1) efektów losowych dla domen $u_{2dt} = \rho_2 u_{2d,t-1} + \epsilon_{2dt}$, gdzie $|\rho_2| < 1$ i $\epsilon_{2dt} \sim N(0, \sigma_2^2)$.
3. Autokorelację przestrzenną SAR(1) efektów losowych dla domen w czasie t $u_{1d} = \rho_1 \sum_{l \neq d} w_{d,l} u_{1,l} + \epsilon_{1d}$, gdzie $|\rho_1| < 1$, $\epsilon_{1d} \sim N(0, \sigma_1^2)$, a $\mathbf{W} = (w_{d,l})$ jest macierzą sąsiedztwa standaryzowaną wierszowo.

EBLUP dla tego modelu ma postać

$$\hat{y}_{dt}^{\text{EBLUP MMM}} = \mathbf{x}'_{dt} \hat{\boldsymbol{\beta}} + \hat{u}_{1d} + \hat{u}_{2dt} \quad (4)$$

gdzie $d = 1, 2, \dots, D$ i $t = 1, 2, \dots, T$.

WYNIKI

Wykorzystano pakiet statystyczny **R** oraz dwa pakiety:

- *sae* oraz funkcje *eblupDyn* i *eblupRy* (Fay i Diallo, 2015),
- *sae2* oraz funkcję *eblupSTFH* (Molina i Marhuenda, 2015).

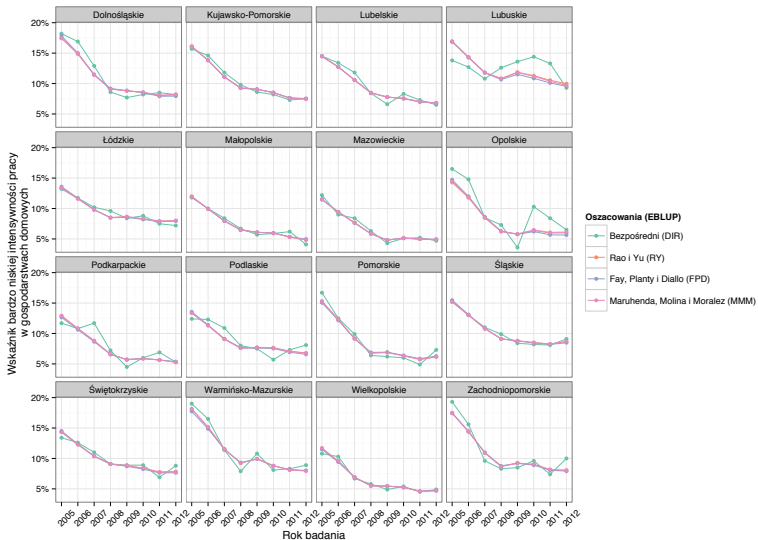
Rozważane zmienne objaśniające:

- Stopa bezrobocia rejestrowanego
- Udział liczby osób w wieku produkcyjnym w ludności ogółem
- Udział liczby osób w wieku produkcyjnym w liczbie osób w wieku poprodukcyjnym

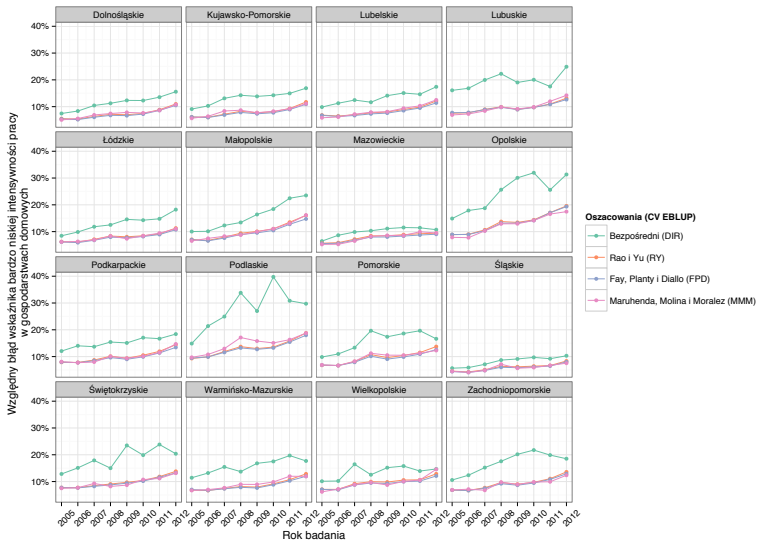
OSZACOWANE MODELE - PORÓWNANIE

	EBLUP RY	EBLUP FPD	EBLUP MMM
Wyraz Wolny	1,2781 (<0,001)	1,4506 (<0,001)	1,2491 (<0,001)
Stopa Bezr Rej	0,0037 (<0,001)	0,0035 (<0,001)	0,0038 (<0,001)
Lud Prod/poprod	0,0894 (<0,001)	0,0923 (<0,001)	0,0907 (<0,001)
Udz Lud Prod	-2,4765 (<0,001)	-2,7668 (<0,001)	-2,4401 (<0,001)
R2	84,17	84,68	84,38
LogLik	335,52	336,76	335,68
BIC	-637,08	-639,56	-632,54
AIC	-657,04	-659,52	-655,36
ρ_1 (SAR(1))	-	-	0,4730
ρ_2 (AR(1))	0,9753	0,9397	0,9082

OSZACOWANIA



WZGLĘDNE BŁĘDY SZACUNKU



WZGLĘDNE BŁĘDY SZACUNKU

Estymator	Min	Mediana	Średnia	Max
Bezpośredni	5,67	14,87	15,81	39,75
Rao i Yu	4,26	8,95	9,34	19,59
Fay, Planty i Diallo	4,09	8,69	9,02	19,34
Marhuenda, Molina i Moralez	4,28	8,70	9,22	19,99

PODSUMOWANIE

- Statystyka małych obszarów może być istotnym narzędziem służącym do otrzymywania odpowiedniej jakości oszacowań, w sytuacjach gdy wariancja estymacji bezpośredniej jest nieakceptowalna
- zastosowanie dynamicznego modelu uwzględniającego autokorelację w czasie spowodowało wyeliminowanie wahań przypadkowych wynikających z doboru próby
- uzyskane oszacowania były w większości przypadków zbieżne, natomiast najmniejszymi wartościami względnych błędów szacunku charakteryzował się model Fay, Planty i Diallo

- Fay, R.E. and Herriot, R.A. (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association* 74, 269-277.
- Fay, R.E., Planty, M. and Diallo, M.S. (2013). Small area estimates from the National Crime Victimization Survey. *Proceedings of the Joint Statistical Meetings. American Statistical Association*, pp. 1544-1557.
- Fay R.E. and Diallo M. (2015). sae2: Small Area Estimation: Time-series Models. R package version 0.1-1. <http://CRAN.R-project.org/package=sae2>
- Marhuenda, Y., Molina, I. and Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis* 58, 308-325.
- Molina I and Marhuenda Y (2015). sae: Small area estimation. R package version 1.0-5, <http://CRAN.R-project.org/package=sae>.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

DZIĘKUJEMY ZA UWAGĘ

ESTYMACJA MSE OMAWIANYCH ESTYMATORÓW

1. Estymator bezpośredni y_{dt}^{dir} - oszacowanie metodą nieparametrycznego bootstrapu odzwierciedlającego schemat losowania.
2. Estymator $y_{dt}^{EBLUP D RY}$ - oszacowany za pomocą formuły "3g" na podstawie oszacowania metodą modelu metodą REML.
3. Estymator $y_{dt}^{EBLUP D FPD}$ - oszacowany za pomocą formuły "3g" na podstawie oszacowania metodą modelu metodą REML.
4. Estymator $y_{dt}^{EBLUP MMM}$ - oszacowany za pomocą parametrycznego bootstrapu generującego efekty losowe z AR(1) oraz SAR(1). Argumenty wejściowe są oszacowane na podstawie REML.