Appendix 1. Specification of the conditional Poisson-Binomial model with cell indices age-category, gender and region

$$Z_{ijk}|\mu_{ijk} \overset{indep}{\sim} \mathcal{P}oisson\left(\mu_{ijk}\right)$$

$$log(\mu_{ijk}) = log(E_{ijk}) + a \ + \ b \cdot X^{(\mu)}_{ijk} \ + \ e_{ijk}$$

where $e_{ijk} \overset{indep}{\sim} \mathcal{N}ormal\left(0, \sigma_e^2\right)$

$$Y_{ijk}|Z_{ijk}, p_{ijk} \overset{indep}{\sim} \mathcal{B}inomial\left(Z_{ijk}\ ,\ p_{ijk}\right)$$

$$logit(p_{ijk}) = \alpha \ + \ \beta \cdot X^{(p)}_{ijk} \ + \ \varepsilon_{ijk}$$

where $\varepsilon_{ijk} \overset{indep}{\sim} \mathcal{N}ormal\left(0, \sigma_\varepsilon^2\right)$ for $i = 1..I\ ,\ j = 1..J\ ,\ k = 1..K$

cell indices are:
i-th age with categories "0-34" , "35-64" , "65-x"
j-th gender, with categories "Male" , "Female"
k-th region, with categories of the 174 statistical sub-regions of Hungary

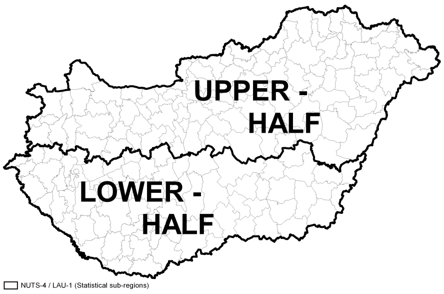
Z_{ijk}, Y_{ijk} are the observed incidences,

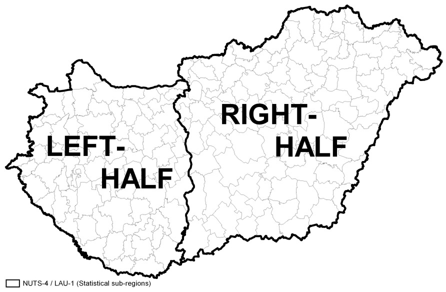
E_{ijk} is the age-gender standardised expected incidence,

$a, b, \sigma_\varepsilon^2, \alpha, \beta, \sigma_e^2$ are unknown parameters,

$X^{(\mu)}, X^{(p)}$ are known explanatory variables.

Appendix 2. An example of artificial data by simulation scheme of two independent halvings





$Z_i \overset{indep}{\sim} Poisson\,(\mu_i)$

for $i = 1, .., 174$ sub-regions

$\mu_i = E_i \cdot b_i$ where $E_i = 0.006 \cdot N_i$

N_i is the population of sub-region i

$$b_i = \begin{cases} 0.9 \text{ in the upper} - \text{half} \\ 1 \text{ in the lower} - \text{half} \end{cases}$$

D1+D2 is a realisation of Z

$Y_i^{(1)} \overset{indep}{\sim} Poisson\left(\mu_i^{(1)}\right)$

for $i = 1, .., 174$ sub-regions,

$\mu_i^{(1)} = \beta_i \cdot \mu_i$ where μ_i

$$\beta_i = \begin{cases} 0.4 \text{ in the right} - \text{half} \\ 0.45 \text{ in the left} - \text{half} \end{cases}$$

D1 is a realisation of $Y^{(1)}$

D2 comes from **D1+D2** and **D1**

Parameter estimation of the conditional Poisson-Binomial model

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
a	0.05775	0.04782	0.002822	-0.03371	0.05752	0.1553	751	2250
b	0.7903	0.1915	0.01496	0.4059	0.8205	1.3557	751	2250
deviance	1884.0	20.08	0.4589	1846.0	1883.0	1924.0	751	2250

Parameter estimation for **D1 + D2**: indicator variable of upper/lower halving of sub-regions are used here and its effect proved to be significant (because coefficient **b** has .95-level confidence interval (0.41 , 1.36) outside zero).

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
a	0.0456	0.04457	0.002497	-0.04171	0.04605	0.1348	751	2250
b	-0.6552	0.09433	0.005195	-0.8395	-0.6529	-0.4698	751	2250
alpha	-0.02203	0.02643	0.001492	-0.07451	-0.02159	0.02898	751	2250
beta	0.2013	0.05598	0.003062	0.09089	0.2025	0.3097	751	2250
deviance	1929.0	16.99	0.9391	1894.0	1930.0	1959.0	751	2250

Parameter estimation for joint model of **(D1+D2)** and **D1/(D1 + D2)**: coefficient **b** stands for upper/lower halving while coefficient **beta** for left/right halving, and both of them proved to be significant. This example demonstrates that the model can handle **(D1+D2)** and **D1/(D1 + D2)** depending on different explanatories.

Appendix 3. Specification and an application of the conditional Poisson-Binomial model with cell indices age-category, gender, sub-region and year

$Z_{ijkt}|\mu_{ijk} \overset{indep}{\sim} \mathcal{P}oisson\left(\mu_{ijkt}\right)$

$log(\mu_{ijkt}) = log(E_{ijkt}) + a \ + \ b \cdot X_{ijk}^{(\mu)} \ + \ e_{ijkt}$

where $e_{ijkt} \overset{indep}{\sim} \mathcal{N}ormal\left(0, \sigma_e^2\right)$

$Y_{ijkt}|Z_{ijkt}, p_{ijk} \overset{indep}{\sim} \mathcal{B}inomial\left(Z_{ijkt} \ , \ p_{ijk}\right)$

$logit(p_{ijk}) = \alpha \ + \ \beta \cdot X_{ijk}^{(p)} \ + \ \varepsilon_{ijk}$

where $\varepsilon_{ijk} \overset{indep}{\sim} \mathcal{N}ormal\left(0, \sigma_\varepsilon^2\right)$ for $i = 1..I$, $j = 1..J$, $k = 1..K$

cell indices are:
i-th age with categories "0-34" , "35-64" , "65-x"
j-th gender, with categories "Male" , "Female"
k-th region, with categories of the 174 statistical sub-regions of Hungary
t-th year, with categories "2004" , "2005" , "2006" , "2007" , "2008"

Z_{ijkt}, Y_{ijkt} are the observed incidences,

E_{ijk} is the age-gender standardised expected incidence,

$a, b, \sigma_\varepsilon^2, \alpha, \beta, \sigma_e^2$ are unknown parameters,

$X^{(\mu)}, X^{(p)}$ are explanatory variables

The following example deals with explanatory variable "Social indicator of the economic underdevelopment status" based on KSH (Hungarian Central Statistical Office) Labor Force Dataset.

Outcome variable Z is the sum of incidence numbers of (K25 + K26) by age-category, gender, sub-region and year, while variable Y is incidence numbers of K25.

Input data can be visualised as follows.

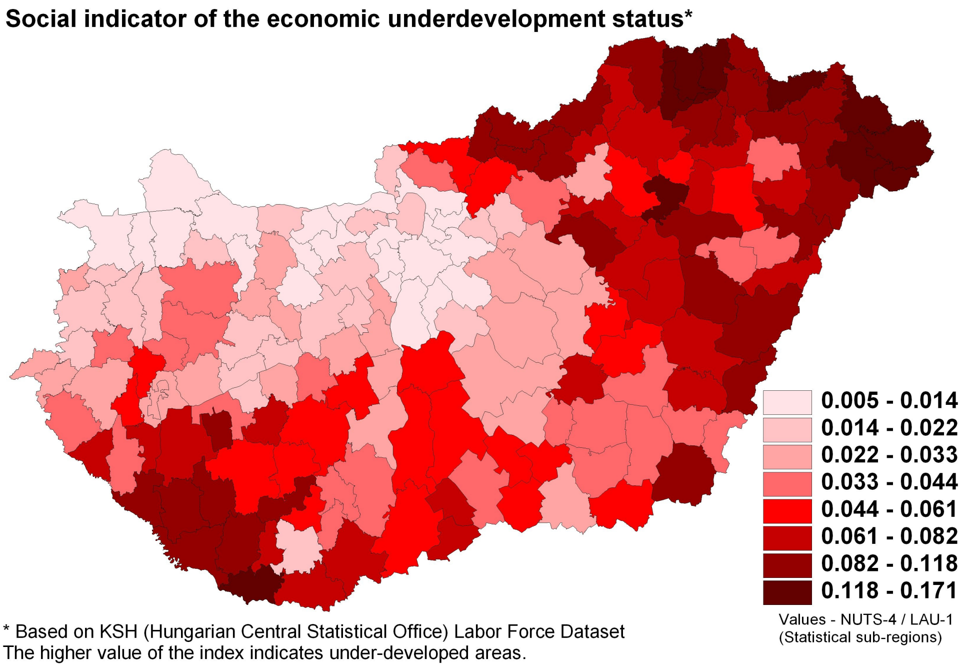


Figure 2. Social indicator of the economic underdevelopment status of 174 sub-regions of Hungary

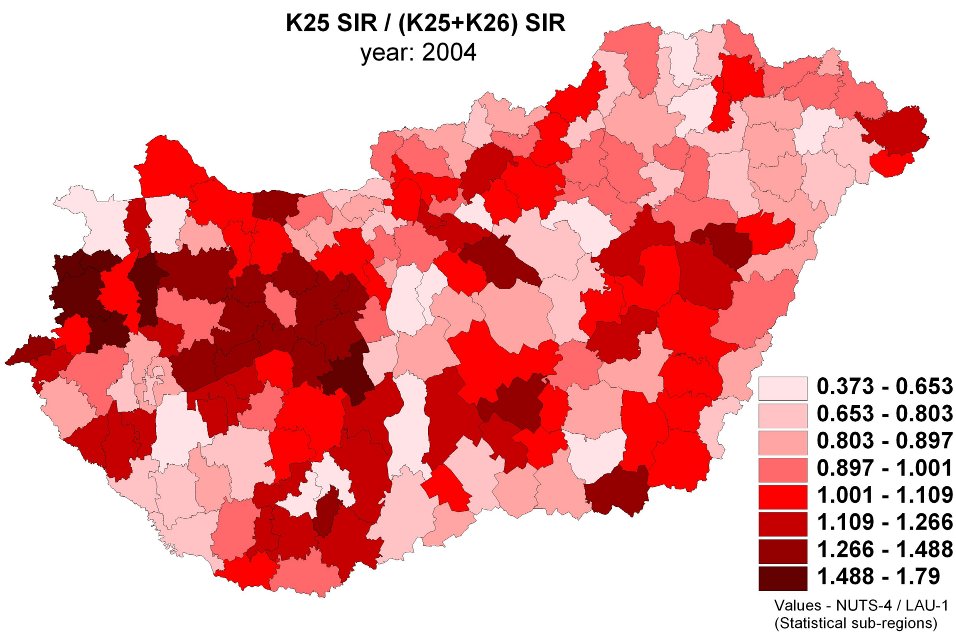


Figure 3. K25 SIR / (K25+K26) SIR values of 174 sub-regions of Hungary, 2004

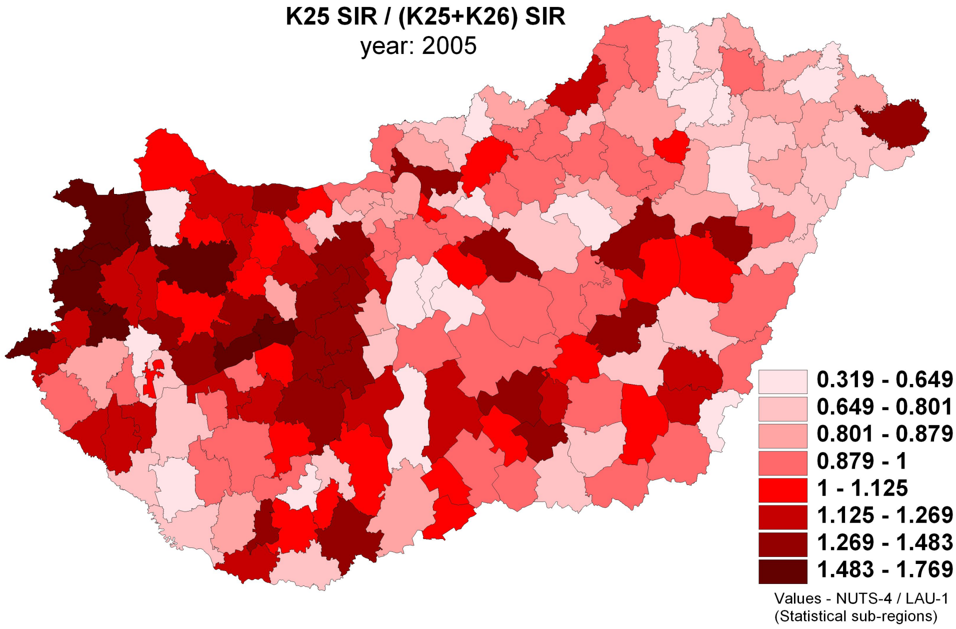


Figure 4. K25 SIR / (K25+K26) SIR values of 174 sub-regions of Hungary, 2005

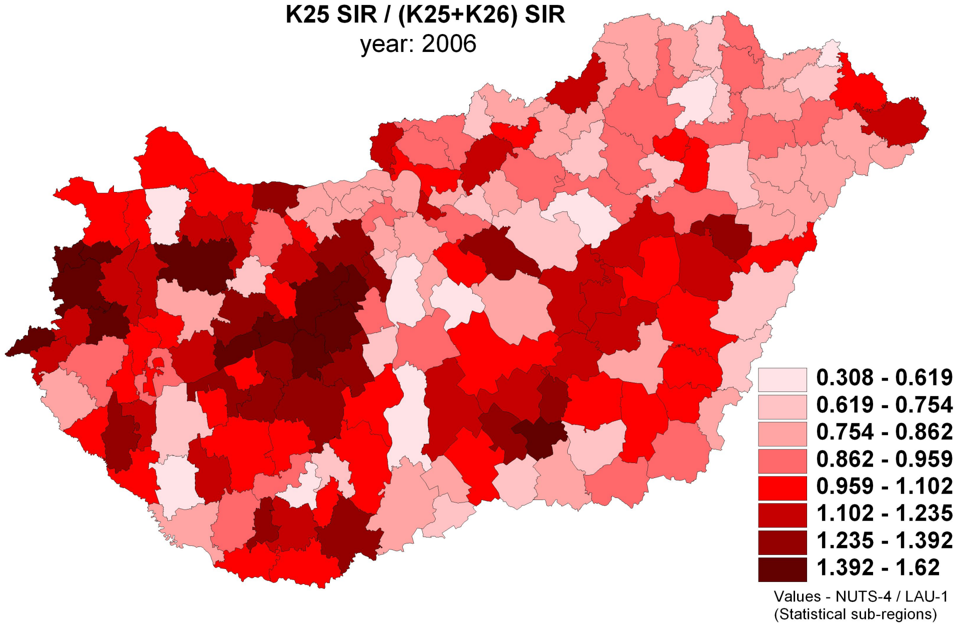


Figure 5. K25 SIR / (K25+K26) SIR values of 174 sub-regions of Hungary, 2006

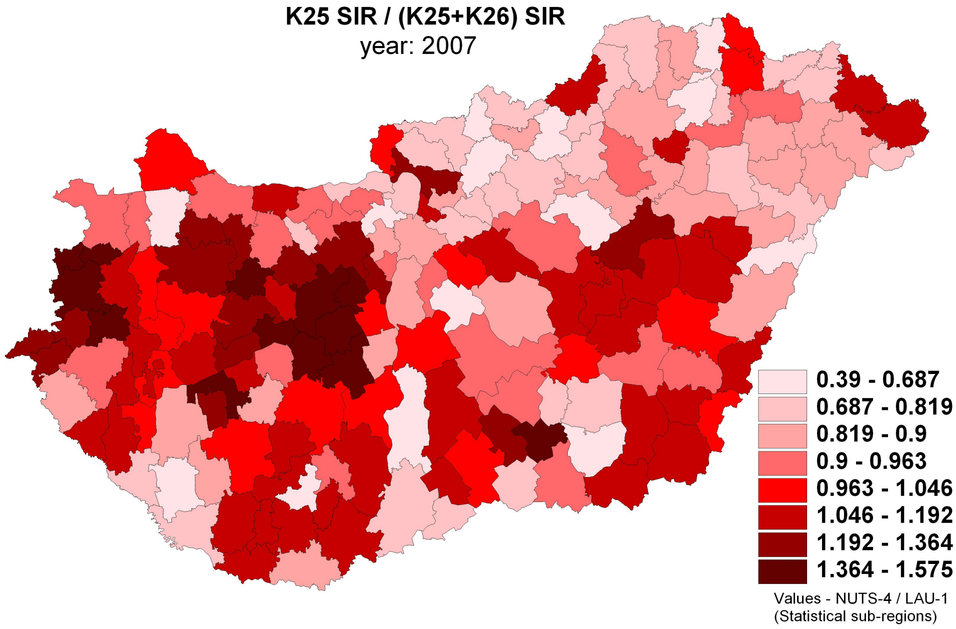


Figure 6. K25 SIR / (K25+K26) SIR values of 174 sub-regions of Hungary, 2007

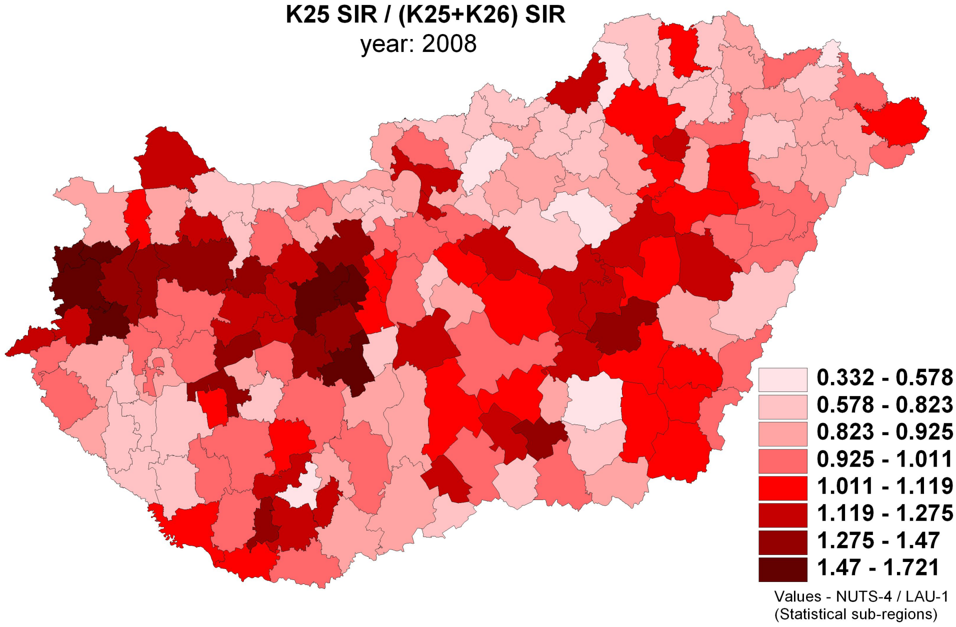


Figure 7. K25 SIR / (K25+K26) SIR values of 174 sub-regions of Hungary, 2008

Table 6. Parameter estimation in Poisson-Binomial model with explanatory variable "Social indicator of the economic underdevelopment status".

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
a	-0.1119	0.03112	0.00236	-0.2077	-0.1231	-0.10672	751	2250
b	0.3166	0.2871	0.01819	-0.2115	0.4074	1.0335	751	2250
alpha	0.3821	0.0794	0.00544	0.1983	0.3164	0.5193	751	2250
beta	-0.5812	0.1772	0.00713	-0.2641	-0.6123	-0.9106	751	2250
deviance	2711	33.41	0.9093	2026	2732	3280	751	2250

One can guess that the economic underdevelopment status (shown in Figure 2.) tends to vary inversely with SIR ratios (shown in Figures 3-7.). Bad economic status of East-North and South-West part of Hungary coincides with high K25 SIR / (K25 + K26) SIR areas. This visual impression is confirmed by Table 6. because coefficient **beta** proved to be negative and statistically significant (its 0.95-level confidence interval is [-0.2641 , -0.9106] and zero is outside of this interval).