Comparative Economic Research, Volume 17, Number 4, 2014
10.2478/cer-2014-0038

EDYTA ŁASZKIEWICZ*, GUANPENG DONG, RICHARD HARRIS*****

The Effect Of Omitted Spatial Effects And Social Dependence In The Modelling Of Household Expenditure For Fruits And Vegetables

Abstract

As is well known, ignoring spatial heterogeneity leads to biased parameter estimates, while omitting the spatial lag of a dependent variable results in biasness and inconsistency (Anselin, 1988). However, the common approach to analysing households' expenditures is to ignore the potential spatial effects and social dependence. In light of this, the aim of this paper is to examine the consequences of omitting the spatial effects as well as social dependence in households' expenditures.

We use the Household Budget Survey microdata for the year 2011 from which we took households' expenditures for fruits and vegetables. The effect of ignoring spatial effects and/or social dependence is analysed using four different models obtained by imposing restrictions on the core parameters of the hierarchical spatial autoregressive model (HSAR). Finally, we estimate the HSAR model to demonstrate the existence of spatial effects and social dependence.

We find the omitted elements of the external environment affect negatively the estimates for other spatial (social) effect parameters. Especially, we notice the overestimation of the random effect variance when the social dependence is omitted and the overestimation of the social interaction effect when the spatial heterogeneity is ignored.

Keywords: *social interaction, consumption behavior, spatial multilevel model*

* University of Lodz, Faculty of Economics and Sociology, Department of Spatial Econometrics
** Ph.D., University of Sheffield, Sheffield Methods Institute
*** Full Professor at the University of Bristol, School of Geographical Sciences

1. Introduction

In the spatial econometrics literature, the negative consequences of ignoring the presence spatial autocorrelation and/or spatial heterogeneity are well known (see e.g. Anselin 1988; Anselin and Griffith 1988). Various methods have been proposed to handle the spatial effects, including: spatial econometric model (e.g. Anselin 1988; LeSage and Pace, 2010), spatially switching regression (Anselin 1990), random coefficient models (Longford 1993) and geographically weighted regression (Fotheringham et al. 2003), among others. Recently, growing attention has been given to the synergy between multilevel and spatial econometric modelling to achieve a spatial multilevel approach (Corrado and Fingleton 2012; Baltagi et al. 2014; Dong and Harris 2014).

Despite above, in many areas of microeconomics studies there is still little interest in the spatial and multilevel approach. One such field is the analysis of consumption behaviour in which the exploration of the hierarchical structure of the microdata as well as the spatial or social dependence has rarely been seen. A few exceptions are the work of Ball et al. (2006) and of Giskes et al. (2006). However, the role of spatial context or the impact of others' decisions on a person's own shopping choices is rarely considered. Because consumers are not separated from each other and live in places that differ by, e.g. the accessibility of the products, so we can expect there to be spatial and social dependence. More general we can say that consumption behaviours are affected by the external environment.

The aim of this paper is to examine the consequences of ignoring the spatial effects and/or social dependence in the analysis of consumption behaviour. We use the microdata from the Household Budget Survey of Poland to explore how misleading conclusions might be drawn when the external environment of the consumption choices is omitted. The hierarchical spatial autoregressive (HSAR) model is applied as well as four additional, misspecified models for comparison. The conclusions from our work are potentially useful as they increase awareness about the merits of using a spatial multilevel approach.

The rest of the paper is organized as follows. In the next section we presented the HSAR model and its variations that result from imposing restrictions on the parameters of the HSAR model. The Bayesian MCMC method of estimation is also described briefly. In section three the characteristics of the data we used is provided. After this, the empirical results for the expenditures model is presented in section four. Finally, the conclusion follows.

2. Method

Our data has the multilevel structure with households at the individual level and statistical survey points at the community level. Due to this, we use the hierarchical spatial autoregressive model (HSAR) proposed by Dong and Harris (2014). The general formula of the model is as follows:

$$\begin{aligned} &\mathbf{Y} = \rho\mathbf{WY} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\Delta\theta} + \boldsymbol{\varepsilon}, \\ &\boldsymbol{\theta} = \lambda\mathbf{M}\boldsymbol{\theta} + \boldsymbol{\mu}, \\ &\varepsilon \sim N(0, \mathbf{I}_N \sigma_\varepsilon^2), \\ &\mu \sim N(0, \mathbf{I}_J \sigma_\mu^2), \end{aligned} \tag{1}$$

where Y is an N×1 vector of a dependent variable. In this research, the dependent variable was specified as the logarithm of the monthly households' expenditures for fruits and vegetables. The N=37 375 is the total number of households in the sample. The X is an N×K matrix of control variables (with constant), while β is a K×1 vector of coefficients to estimate. The N×1 vector of error terms was assigned as ε, while μ is J×1 vector of random effects for the communities. The total number of communities is J=1 551. We assumed that both the error term and random effects are normally distributed with variance σε2 and σμ2, respectively. The estimated parameter of the social interaction between households is ρ and λ is the estimated parameter of the spatial interaction between communities. The N×J block-diagonal design matrix Δ is as follows:

$$\boldsymbol{\Delta} = \begin{bmatrix} \mathbf{l}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{l}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{l}_J \end{bmatrix}, \tag{2}$$

where: 0 is nj×1 vector of zeroes and nj is the number of households located in the community j.

In the N×N social interaction matrix W we specified the structure of the relationships between each pair of households. It can be written as:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_J \end{bmatrix}. \tag{3}$$

The elements, w_{ij} of each submatrices $\mathbf{W}_j$ were calculated as the inverse exponential function of the time differences in months. The time means the month when the household declared the expenditures for fruits and vegetables. The result of applying above weights is the assignation of the higher weights (the stronger relationship between households) for those pairs which declared the expenditures in the same or adjacent time. The greater the difference between the declaration about the expenditures, the weaker is the potential influence in a pair of the households. It can be written as:

$$w_{ij} = \begin{cases} 1/\exp(\Delta t)^2 \text{ if } t_i \geq t_{i'} \wedge i \neq i' \\ 0 \text{ if } t_i < t_{i'} \vee i = i' \end{cases}, \tag{4}$$

where: $\Delta t = t_i - t_{i'}$, $t_i = 1, \ldots, 12$ denotes the time when household i declared expenditures for fruits and vegetables.

Moreover, the $J \times J$ block-diagonal spatial matrix $\mathbf{M}$ was specified to capture the spatial interactions between communities located in the same voivodship (region). It is as follows:

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_R \end{bmatrix}. \tag{5}$$

The elements of each from the R=16 submatrices $\mathbf{M}_r$ were specified as a binary function with the value one if two communities are located in the same voivodship, or zero otherwise. Only the group-wise spatial dependence was applicable to spatial relationships between communities as only the information about the community location in the voivodship was known. Using the notation in Corrado and Fingleton (2012) it can be expressed as:

$$\mathbf{M}_r = \frac{1}{n_r}\left(\mathbf{l}_{nr}\mathbf{l}_{nr}^{'}\right), \tag{6}$$

where: $\boldsymbol{l}_{nr}$ is a $n_r \times 1$ column vector of ones**.** Both **W** and **M** matrices were row-standardized.

Additional models, which are used in the research, are obtained by imposing the restrictions on the HSAR parameters. Four different scenarios were studied: lack of both spatial effects and social dependence, omitted spatial effects, lack of social dependence and lack of spatial dependence. They were achieved as follows:

- λ=0 and σ_μ^2=0, which is the equivalent of a standard spatial autoregressive model. In this model we ignore the spatial effects. It means there is no spatial interaction between communities (λ=0) as well as no differences between communities in the level of expenditures for fruits and vegetables (σ_μ^2=0);
- ρ=0 and λ=0, which gives a standard multilevel model in which we ignore the potential social interactions between households (ρ=0) and spatial interactions between communities (λ=0);
- ρ=0, which means we allow for the spatial effects but omit the potential social dependence. In this model we concentrate only on the heterogeneity and dependence effects at the community level, and
- λ=0, which is the equivalent of a model with social interactions at the individual level and spatial heterogeneity at the community level. This model was achieved by combing the spatial autoregressive model and multilevel model. Although it allows for the inequalities of the expenditures at the community level, the potential spatial interactions among communities is not modelled.

The Bayesian Markov Chain Monte Carlo (MCMC) method was used to estimate the HSAR and four additional models. According to Dong and Harris (2014) the prior distributions for each parameter in the HSAR model are specified as follows:

$$\begin{aligned} &P(\boldsymbol{\beta}) \sim N(\mathbf{M}_0, \mathbf{T}_0), \\ &P(\rho) \sim U(1/\upsilon_{\rho\min}, 1),\ P(\lambda) \sim U(1/\upsilon_{\lambda\min}, 1), \\ &P(\sigma_\varepsilon^2) \sim IG(c_0, d_0),\ P(\sigma_\mu^2) \sim IG(a_0, b_0), \end{aligned} \tag{7}$$

where: $\mathbf{M}_0$ is the K×1 vector of means, $\mathbf{T}_0$ is the variance matrix, υ_{min} is the minimum eigenvalue of the weight matrix, IG is the inverse gamma distribution with the shape parameter a_0 or c_0 and scale parameter b_0 or d_0, N is the normal distribution and U is the uniform distribution.

The full posterior conditional distributions for each model parameter are derived based on the likelihood function for the HSAR model and the prior distributions (see Dong and Harris, 2014). Hence, the conditional posterior distributions were:

$$P(\boldsymbol{\beta} \mid \mathbf{Y}, \rho, \lambda, \boldsymbol{\theta}, \sigma_\varepsilon^2, \sigma_\mu^2) \sim N(\mathbf{M}_\beta, \boldsymbol{\Sigma}_\beta), \tag{8}$$

with:

$$\begin{aligned} \boldsymbol{\Sigma}_\beta &= \left[\left(\sigma_\varepsilon^2\right)^{-1}\mathbf{X}'\mathbf{X} + \mathbf{T}_0^{-1}\right]^{-1}, \\ \mathbf{M}_\beta &= \boldsymbol{\Sigma}_\beta\left[\left(\sigma_\varepsilon^2\right)^{-1}\mathbf{X}'\left(\mathbf{AY} - \Delta\boldsymbol{\theta}\right) + \mathbf{T}_0^{-1}\mathbf{M}_0\right] \end{aligned} \tag{9}$$

$$P(\boldsymbol{\theta} \mid \mathbf{Y}, \rho, \lambda, \boldsymbol{\beta}, \sigma_\varepsilon^2, \sigma_\mu^2) \sim N(\mathrm{M}_{\boldsymbol{\theta}}, \Sigma_{\boldsymbol{\theta}}), \tag{10}$$

with:

$$\begin{aligned} \Sigma_\theta &= \left[\left(\sigma_\varepsilon^2\right)^{-1}\Delta'\Delta + \left(\sigma_\mu^2\right)^{-1}\mathbf{B}'\mathbf{B}\right]^{-1}, \\ \mathrm{M}_{\boldsymbol{\theta}} &= \Sigma_\theta\left[\left(\sigma_\varepsilon^2\right)^{-1}\Delta'\left(\mathbf{AY} - \mathbf{X}\boldsymbol{\beta}\right)\right] \end{aligned} \tag{11}$$

$$P(\sigma_\mu^2 \mid \mathbf{Y}, \rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\varepsilon^2) \sim IV(a_\mu, b_\mu), \tag{12}$$

with:

$$\begin{aligned} a_\mu &= J/2 + a_0, \\ b_\mu &= \boldsymbol{\theta}'\mathbf{B}'\mathbf{B}\boldsymbol{\theta}/2 + b_0. \end{aligned} \tag{13}$$

$$P(\sigma_\varepsilon^2 \mid \mathbf{Y}, \rho, \lambda, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma_\mu^2) \sim IV(c_\varepsilon, d_\varepsilon), \tag{14}$$

with:

$$\begin{aligned} c_\varepsilon &= N/2 + c_0, \\ d_\varepsilon &= 0{,}5 \times \left(\mathbf{AY} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta\theta}\right)'\left(\mathbf{AY} - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\Delta\theta}\right) + d_0. \end{aligned} \tag{15}$$

where: $\mathbf{A} = \mathbf{I}_N - \rho\mathbf{W}$ and $\mathbf{B} = \mathbf{I}_J - \lambda\mathbf{M}$. The Gibbs sampler was employed to draw the samples for parameters.

Because the posterior distributions for ρ and λ do not fit standard recognizable density distributions, the inverse sampling method was used to update the social and spatial interaction parameters. More specifically, in each iteration after the numerical integration of Log $f(\rho)$ over $(1/\upsilon_{\rho min},1)$ and Log $f(\lambda)$

over ($1/\upsilon_{\lambda\min}$, 1), the cumulative distribution of ρ and λ were calculated. Then, the inverse sampling approach was employed to draw values of both parameters. The Log $f(\rho)$ and Log $f(\lambda)$ in the HSAR model are as follows:

- For the social interaction parameter ρ:

$$\text{Log } f(\rho) = \log\left|\mathbf{I}_N - \rho\mathbf{W}\right| + S(\rho)' S(\rho) / 2\sigma_\varepsilon^2 + m, \quad (16)$$

with:

$$\begin{aligned} S(\rho) &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}_0) - \rho(\mathbf{WY} - \mathbf{X}\boldsymbol{\beta}_d) - (\boldsymbol{\Delta\theta} - \mathbf{X}\boldsymbol{\beta}_u), \\ \boldsymbol{\beta}_0 &= (\mathbf{X'X})^{-1}\mathbf{X'Y}, \\ \boldsymbol{\beta}_d &= (\mathbf{X'X})^{-1}\mathbf{X'WY}, \\ \boldsymbol{\beta}_u &= (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\Delta\theta}. \end{aligned} \quad (17)$$

- For the spatial interaction parameter λ:

$$\text{Log } f(\lambda) = \log\left|\mathbf{I}_J - \lambda\mathbf{M}\right| + \boldsymbol{\theta}'\mathbf{B}'\mathbf{B}\boldsymbol{\theta} / 2\sigma_\mu^2, \quad (18)$$

where: m is a constant.

For further discussion about the MCMC algorithm for implementing the HSAR model see Dong and Harris (2014). The MCMC samplers for the HSAR model and other four models were coded using the R language. The convergence of the MCMC samplers was diagnosed using the CODA package in R Cran (Plummer et al. 2006). The inferences were based on one MCMC chain that each consist of 10 000 iterations with a burn-in period of 5000 for each model. For all models, diffuse or quite non-informative priors are used for parameters while the initial values are drawn randomly from their prior distributions.

3. Database and descriptive statistics for dependent and control variables

The microdata used in this study comes from the 2011 Polish Household Budget Survey. It is the largest and most representative survey for household expenditure in Poland, conducted by the Central Statistical Office (GUS). The full sample consists of N=37 375 households as each of these households declared non-zero expenditures for fruits and vegetables. The community level was defined as the area survey point and consists of J=1 551 spatial units. The category of fruits and vegetables expenditures was separated with consistency with the Classification of Expenditures on Consumer Goods and Services (GUS, 2011, pp. 256-257).

In 2011, the average monthly households' expenditure for fruits and vegetables was 118 PLN which accounted for almost 18% of the total food expenditures. The distribution of the expenditures was characterized by the high, positive kurtosis (11,69) and skewness (1,97). Hence, to approximate the normal distribution, the log transformation was used for the value of households' expenditures and the transformed variable was taken as the dependent variable in our models.

As for the control variables, we used those which represent households' socio-economic status and personal attributes of the reference person. The household profile was characterized by the number of persons in the household (h_size) and by the type of the household. The type was classified into four dummy-variables: couples with children (couple_ch), couples without a child (couple_nch), singles (single) and others (reference category). The mean values of expenditures from each type of household were found to be significantly different from each other. The highest expenditures for fruits and vegetables were noticed for couples with children (138 PLN). To control for the household size effect we repeated the analysis using the value of expenditures per capita but the differences between types of household was still significant. The logarithm of the households' monthly available income was used to capture the economic conditions of the household. To avoid potential multicollinearity only two personal attributes were taken into account: sex (1 if male) and age (as a continuous variable) of the reference person.

In addition, we found the value of expenditures for fruits and vegetables varies between the hierarchies of locality. As shown in Table 1, median expenditure per capita was significantly higher in the biggest Polish cities with the population over 500 thousand (ref.category) than in the other cities and rural areas in 2011. The median decreases with the city size and the lowest was noticed for villages (34,26 PLN). Hence, we added five dummy-variables to represent the hierarchy of locality: cities with 200-499 thous. inhabitants (cities_1), towns with population 100-199 thous. (cities_2), towns with 20-99 thous. (cities_3), towns with less than 20 thous. inhabitants (cities_4) and rural areas (rural).

Table 1. Descriptive statistics for households' expenditures for class of locality

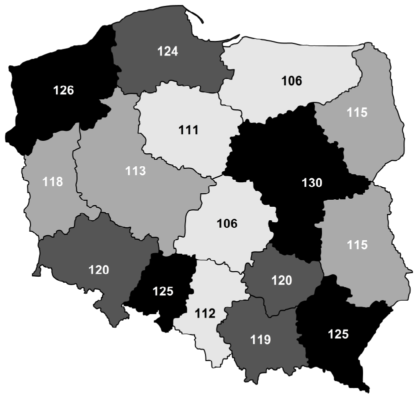
class of locality	expenditures per capita*		
	median	quantile 25%	quantile 75%
ref. category	50,28	32,22	77,33
cities_1	43,65	28,36	66,60
cities_2	41,31	26,17	63,50
cities_3	40,76	25,99	63,69
cities_4	38,58	24,41	59,36
rural	34,26	22,66	52,40

* In Polish zloty.

Source: authors' own.

An additional source of the place-related variation in expenditure was due to the household location in the voivodship (see Figure 1). According to the results from the Wilks' lambda test for equality of 16 group means (conducted on both the value of expenditures and the value per capita), there are regional differences in the level of fruits and vegetables consumption. The regional differences were captured by allowing for the inter-regional dependence in the spatial matrix **M**. Another way to incorporate it would be adding the regional fixed effects.

Figure 1. Average household expenditures for fruits and vegetables in Polish voivodships



Source: authors' own.

4. Empirical results

We started our analyses from the results for the OLS and SAR models (Section 4.1) to check for the existence of the social dependence. Then we estimated the MLM model (Section 4.2) to find the spatial heterogeneity at the community level. In the next two subsections, the results for the HSAR models with ρ=0 (Section 4.3) and λ=0 (Section 4.4) were analyzed, while in the last one (Section 4.5) the results for the HSAR model without restrictions on parameters were presented.

4.1. Omitted spatial heterogeneity and dependence

In the OLS model, 22% of the total variance of households' expenditures for fruits and vegetables were explained[1]. The signs of the estimated parameters for all control variables were as expected (see Table 2). We computed Moran's I test for the OLS residuals using the social interaction matrix W. The Moran's I statistic was equal to 0,1 and significant at the 0,01 level. Because the OLS residuals are correlated, we should incorporate the social dependence into our model to avoid the misspecification.

Following the sequence of models outlined above, we first estimated a SAR model, which is the equivalent of the HSAR model with the following restrictions on parameters: λ=0, $\sigma\mu2$=0. It means we allow only for the social dependence effect while leave unexamined spatial heterogeneity and dependence effects at the community level. The results are shown in Table 2. The estimated error variance was lower than in the OLS model by about 2,4%. The 95% credible intervals for the $\sigma\varepsilon2$ in the SAR model do not contain the value of $\sigma\varepsilon2$ from the OLS. This might suggest an overestimation of the error variance in the OLS.

Table 2. Estimation results for fruits and vegetables expenditures using OLS and SAR models

variable	OLS		SAR			
	coef.	std.err.	posterior mean	std. error	2,5%	97,5%
intercept	3,261	0,036	2,506	0,044	2,476	2,592
log_income	0,123	0,004	0,118	0,004	0,116	0,125
h_size	0,107	0,004	0,107	0,004	0,105	0,114
age	0,004	0,000	0,004	0,000	0,004	0,004
sex	-0,025	0,007	-0,027	0,007	-0,032	-0,013
couple_ch	0,066	0,011	0,068	0,011	0,060	0,088
couple_nch	0,038	0,010	0,040	0,010	0,034	0,059
single	-0,411	0,013	-0,406	0,013	-0,415	-0,381
cities_1	-0,079	0,014	-0,071	0,014	-0,081	-0,045
cities_2	-0,110	0,015	-0,104	0,015	-0,113	-0,075
cities_3	-0,103	0,012	-0,098	0,012	-0,106	-0,075
cities_4	-0,141	0,013	-0,132	0,013	-0,141	-0,107
rural	-0,129	0,011	-0,132	0,011	-0,139	-0,111
ρ			0,174	0,007	0,170	0,187
σ_ε^2	0,379		0,370	0,003	0,368	0,375

Source: Own calculations in R Cran.

[1] It is typical to obtain the low value of the R-squared for the models based on the micro data (see e.g. Cameron, Trivedi, 2005, p. 7).

We find that the estimate for the social interaction (ρ) is positive and significant. It suggests that interpersonal relationships affect the level of household consumption of fruits and vegetables. Although we allowed for the social dependence, the estimation for the control variables did not change significantly.

4.2. Omitted social and spatial dependence

Next we estimated the multilevel model (MLM) obtained by adding the restrictions on the HSAR parameters: $\lambda=0$ and $\rho=0$. In the MLM model only spatial heterogeneity was allowed and we assumed no social or spatial dependence. The spatial heterogeneity was defined as the difference in the households' expenditures between communities. The estimation results are presented in Table 3.

Table 3. Estimation results for fruits and vegetables expenditures using a MLM model

variable	posterior mean	std. error	2,5%	97,5%
intercept	3,320	0,034	3,316	3,413
log_income	0,113	0,004	0,110	0,120
h_size	0,108	0,004	0,106	0,116
age	0,004	0,000	0,004	0,004
sex	-0,029	0,007	-0,033	-0,015
couple_ch	0,067	0,011	0,060	0,088
couple_nch	0,043	0,010	0,036	0,061
single	-0,410	0,013	-0,418	-0,385
cities_1	-0,081	0,002	-0,096	-0,036
cities_2	-0,112	0,024	-0,129	-0,065
cities_3	-0,105	0,019	-0,120	-0,067
cities_4	-0,143	0,021	-0,158	-0,102
rural	-0,134	0,017	-0,146	-0,101
σ_{μ}^2	0,028	0,002	0,027	0,031
σ_{ε}^2	0,351	0,003	0,350	0,356

Source: Own calculations in R Cran.

We notice the significant decrease of the error variance in comparison to the OLS model (7,4%) as well as the SAR (5,1%). The random effect variance is significant, which suggests the existence of spatial heterogeneity at the community level. Such variation of the households' expenditures was 8,0% of the total variance. Additionally, and as with to the results from the SAR model we find no significant changes in terms of estimation for the control variables.

Again, only the estimate for the intercept seems to be affected by the omitted social dependence (in the SAR model) and spatial heterogeneity (in the MLM model). Although the estimation results for fruits and vegetables expenditures from the MLM model look very convincing, we should check for the social dependence, which is suggested from the estimation results in the SAR model.

We checked if the residuals from the MLM model are correlated using Moran's I test. The statistic was insignificant proving lack of the social autocorrelation in the residuals. The same procedure was applied to test the presence of spatial dependence in the community random effects. We separated the estimated random effects as $\mathbf{\Delta\theta}$ and used the matrix **W** to conduct the Moran's I test. The results support the hypothesis of the existence of the social dependence. The value of the Moran's I statistic was equal to 0,1 with a 0,01 significance level. However, as long as the structure of the interaction in matrix **W** is based on the inter-community relations such result for $\mathbf{\Delta\theta}$ seems to be obvious.

We also test the estimated random effects for the presence of spatial dependence using matrix **M**. The Moran I statistic was equal to 0,6 and significant at the 0,01 level. It suggests that the assumption that the community specific effects are independent has been violated. We expected in this case that the estimated random effect variance was affected not only by the omitted social dependence but also by the additional spatial dependence. It might result in an overestimation of the variance but further research is necessary to answer the question about how the spatial and social dependence affect the estimates for the random effect variance.

4.3. Omitted social dependence

In the next step, we estimate the model with both spatial heterogeneity and spatial dependence at the community level but without social dependence between households. The model we achieve was the equivalent of the HSAR model with ρ=0. We estimated it because we are interested in the nature of model misspecification connected with the omitted social interactions.

The estimation results for the HSAR model without social dependence are presented in Table 4. Again, the estimates for the control variables were similar in terms of statistical inference, when compared to those obtained from the OLS, SAR and MLM models. The value of the estimated intercept was not significantly different from that obtained by using the MLM model but was significantly higher than that in the OLS and SAR models.

The estimate for the parameter λ (the measure of spatial dependence) was significant, which suggests the necessity of incorporating it in the model. As long as we allowed for the spatial interactions between communities the estimated value of the random effect variance decreased significantly (about 10,7%). The results obtained support the observation that the random effect variance is overestimated when the spatial dependence is omitted. Despite this, in the HSAR model with ρ=0 we might expect that both the random effects, the error variance and the estimate for λ are biased because of the existence of social dependence among households.

Table 4. Estimation results for fruits and vegetables expenditures using HSAR model with ρ=0

variable	posterior mean	std. error	2,5%	97,5%
intercept	3,331	0,042	3,303	3,414
log_income	0,113	0,004	0,110	0,120
h_size	0,109	0,004	0,106	0,116
age	0,004	0,000	0,004	0,004
sex	-0,027	0,008	-0,033	-0,013
couple_ch	0,067	0,011	0,060	0,088
couple_nch	0,043	0,010	0,036	0,062
single	-0,411	0,013	-0,420	-0,385
cities_1	-0,073	0,025	-0,090	-0,025
cities_2	-0,089	0,027	-0,106	-0,037
cities_3	-0,091	0,020	-0,106	-0,053
cities_4	-0,137	0,021	-0,151	-0,094
rural	-0,123	0,018	-0,137	-0,091
λ	0,709	0,056	0,674	0,808
σ_μ^2	0,025	0,002	0,023	0,028
σ_ε^2	0,386	0,003	0,384	0,392

Source: Own calculations in R Cran.

4.4. Omitted spatial dependence

In contrast to the previous model we allow now for the social dependence as well as spatial heterogeneity but omitted the spatial dependence at the community level. We are used the HSAR model with the restriction of λ=0. The estimation result was presented in Table 5.

Again, the estimated parameters for the control variable are stable. That might suggest that as long as the control variables are not correlated with the spatial dependence or heterogeneity and social dependence parameters, there are negligible negative effects on the estimated parameters of such variables due to the omitted ρ, λ or σμ2.

As with the previous models, the omitted spatial or social dependence affected mostly the estimates for the random effect variance. We notice that when we allowed for the social dependence the estimated variance of the random effects decreased significantly (by 46,4% in comparison with the MLM model). Also, we observe a significant decrease of the estimates for ρ, when the spatial heterogeneity was captured (by 15,5% compared to the SAR model). It suggests that the omitted spatial heterogeneity results in an overestimation of the social interaction parameter in the SAR model.

Table 5. Estimation results for fruits and vegetables expenditures using HSAR model with λ=0

variable	posterior mean	std. error	2,5%	97,5%
intercept	2,668	0,048	2,636	2,762
log_income	0,113	0,004	0,111	0,121
h_size	0,108	0,004	0,106	0,115
age	0,004	0,000	0,004	0,004
sex	-0,029	0,007	-0,033	-0,014
couple_ch	0,068	0,011	0,061	0,089
couple_nch	0,043	0,010	0,036	0,062
single	-0,408	0,013	-0,417	-0,383
cities_1	-0,075	0,019	-0,089	-0,037
cities_2	-0,107	0,021	-0,121	-0,065
cities_3	-0,101	0,016	-0,113	-0,070
cities_4	-0,136	0,018	-0,149	-0,102
rural	-0,137	0,015	-0,146	-0,108
ρ	0,147	0,001	0,142	0,160
σ_{μ}^{2}	0,015	0,001	0,013	0,018
σ_{ε}^{2}	0,375	0,003	0,373	0,381

Source: Own calculations in R Cran.

The overestimation was also found for the random effect variance when the social or spatial dependence is not taken into account. The estimated error variance was not significantly different from that in the SAR model but was higher than that from the MLM model (by 6,8%) and lower than that from the HSAR model with ρ=0 (by 2,8%).

4.5. Social dependence, spatial dependence and heterogeneity in the HSAR model

As the final as full model we estimate the HSAR model allowing for both spatial effects and social dependence. According to the estimation results (in Table 6) all of the mentioned above effects were found to be significant for the fruits and vegetables expenditures. The estimate for the error variance from the HSAR model is lower than that from previous models, except for the MLM model. The estimated ρ parameter decreases sharply in comparison with both the SAR and HSAR model with λ=0. This suggests that both models might overestimate the value of ρ. The overestimation can be due to and reflective of the omitted spatial heterogeneity and/or dependence. In contrast, the estimates for the λ parameter do not change significantly in comparison with the HSAR model with ρ=0. It implies that ignoring the social dependence does not significantly affect the estimates for the spatial dependence.

Table 6. Estimation results for fruits and vegetables expenditures using HSAR model

variable	posterior mean	std. error	2,5%	97,5%
intercept	2,980	0,051	2,945	3,082
log_income	0,112	0,004	0,110	0,119
h_size	0,109	0,004	0,106	0,116
age	0,004	0,000	0,004	0,004
sex	-0,028	0,007	-0,032	-0,014
couple_ch	0,067	0,011	0,060	0,088
couple_nch	0,044	0,010	0,036	0,061
single	-0,410	0,013	-0,419	-0,385
cities_1	-0,069	0,023	-0,085	-0,023
cities_2	-0,087	0,024	-0,103	-0,039
cities_3	-0,089	0,018	-0,101	-0,054
cities_4	-0,133	0,020	-0,146	-0,093
rural	-0,127	0,017	-0,138	-0,094
ρ	0,077	0,008	0,072	0,092
λ	0,717	0,055	0,679	0,815
σ_μ^2	0,020	0,001	0,019	0,022
σ_ε^2	0,352	0,003	0,351	0,358

Source: Own calculations in R Cran.

Finally, we find the significant decreased of the estimates for the random effect variance (compare with the results from the Table 3 and 4). The estimate for the random effect variance in this model decreased by 40% and 25%, respectively when compared to the MLM model and the HSAR model with $\rho=0$. The decline suggests the importance of accounting for the spatial and/or social dependence effect when analysing households' expenditure attributes.

5. Conclusions

In this paper we discussed the consequence of ignoring the spatial effects and/or social dependence in the analysis of households' expenditures for fruits and vegetables. We illustrated that the omitted elements of external environment negatively affect the estimates for some important parameters and as a result misleading conclusions might be drawn. According to the results for the households' expenditures, the omitted spatial effects affect the estimation results – for example an overestimation of the social interaction parameter. Analogously, in the presence of social dependence, omitted interpersonal relationships affect results in the overestimation of the random effect variance. The negative consequences are also noticeable in the case of lack of the spatial dependence at the higher level. If the communities are spatially correlated but this correlation is ignored in the model, both the estimates for the parameter of social dependence and spatial heterogeneity are affected. The estimated parameters for the control variables (except the intercept) were found as the least susceptible for the omitted spatial effects and/or social dependence.

References

Anselin L. (1988), *Spatial Econometrics: Methods and Models*, Vol. 4. Springer.

Anselin L. (1990), *Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis*, 'Journal of Regional Science 30.2', pp. 185-207.

Anselin L., Griffith D. A. (1988), *Do spatial effects really matter in regression analysis?*, 'Papers in Regional Science', 65(1), pp. 11-34.

Ball K., Crawford D., Mishra, G. (2006), *Socio-economic inequalities in women's fruit and vegetable intakes: a multilevel study of individual, social and environmental mediators*, 'Public health nutrition', 9(05), pp. 23-630.

Baltagi B. H., Fingleton B., Pirotte A. (2014), *Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices*, 'Journal of Urban Economics', 80, pp. 76-86.

Cameron A.C., Trivedi P.K. (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press.

Central Statistical Office (2012), Household Budget Survey in 2011, Statistical Information and Elaborations, Warsaw.

Corrado L., Fingleton B. (2012), *Where is the economics in spatial econometrics?*, 'Journal of Regional Science', 52(2), pp. 210-239.

Dong, G., Harris, R J. (2014), *Spatial Autoregressive Models for Geographically Hierarchical Data Structures*, Geographical Analysis.

Fotheringham S.A., Brunsdon C., Charlton M. (2003), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, John Wiley & Sons.

Giskes K., Turrell G., Van Lenthe F. J., Brug J., Mackenbach J. P. (2006), *A multilevel study of socio-economic inequalities in food choice behaviour and dietary intake among the Dutch population: the GLOBE study*, 'Public health nutrition', 9(01), pp. 75-83

LeSage J., Pace R. K. (2010), *Introduction to spatial econometrics*, CRC press.

Longford N.T. (1993), *Random Coefficient Models*, Clarendon Press.

Plummer M., Best N., Cowles K., Vines K. (2006), CODA: *Convergence diagnosis and output analysis for MCMC*, R news, 6(1), 7-11.

Acknowledgements: This work was financial supported by grants ETIUDA1 from the Polish National Centre for Science, UMO-2013/08/T/HS4/00512.

Streszczenie

SKUTKI POMINIĘCIA EFEKTÓW PRZESTRZENNYCH I SPOŁECZNYCH ZALEŻNOŚCI W MODELU WYDATKÓW GOSPODARSTW DOMOWYCH NA OWOCE I WARZYWA

Pominięcie przestrzennej heterogeniczności w modelu ekonometrycznym skutkuje błędnym oszacowaniem parametrów, zaś brak uwzględnienia opóźnionej przestrzennie zmiennej zależnej skutkuje obciążeniem i brakiem zgodności estymatora (Anselin 1988). Mimo tego w analizach wydatków gospodarstw domowych efekty przestrzenne oraz interakcje społeczne są najczęściej pomijane.

W pracy skoncentrowano się na skutkach pominięcia efektów przestrzennych i ww. interakcji. W badaniu wykorzystano mikrodane pochodzące z Badania Budżetów Gospodarstw Domowych (2011 r.), dotyczące wydatków na owoce i warzywa. Skutki pominięcia efektów *przestrzennych i/lub interakcji międzyludzkich zweryfikowano wykorzystując hierarchiczny model autoregresji przestrzennej (HSAR) oraz cztery modele uzyskane poprzez nałożenie restrykcji na parametry modelu HSAR.*

Uzyskane wyniki potwierdziły negatywny wpływ pominięcia składowych środowiska zewnętrznego na oszacowania wybranych parametrów. Zaobserwowano przeszacowanie parametru odzwierciedlającego skalę przestrzennej heterogeniczności w sytuacji pominięcia interakcji międzyludzkich oraz przeszacowanie parametru tychże interakcji w sytuacji pominięcia przestrzennej niejednorodności zjawiska.

Słowa kluczowe: *interakcje społeczne, zachowania konsumpcyjne, przestrzenne modele wielopoziomowe*