Is non-response contagious?

The use of spatial information for predicting and weighting survey
nonresponse

Christoph Zangger*

Patterns of nonresponse are often geographically correlated. Existing research has recognized this spatial dependence, using tools to map nonresponse or by means of nonresponse adjustments that draw on geographically aggregated auxiliary information to predict people's participation in surveys. Against this background, this contribution asks whether spatial econometric models can improve nonresponse adjustments and if they provide better propensity weights. Spatial models bear the advantage of an explicit modeling of the underlying spatial process by, for example, including the mutual dependence of neighboring units' response status. Results from Monte Carlo simulations demonstrate that they can indeed yield better estimates of population characteristics, especially if extreme weights are trimmed and when aggregated population shares of unobserved, spatially correlated confounders are available.

Keywords: Nonresponse, Weighting, Spatial Econometrics, Monte Carlo Simulation

1. Introduction

It has widely been noted that some social groups are more likely to participate in social surveys than others (Abraham, Maitland, and Bianchi, 2006; Etter and Perneger,

---

*University of Zurich, zangger@soziologie.uzh.ch; Prepared for the ESA RN 21 mid-term conference 2020 (Original title: "The geography of nonresponse")

1997; Watanabe, Olson, and Falci, 2017). Among other characteristics, response rates differ by invited participants' age (Macera et al., 1990), gender (Criqui, Barrett-Connor, and Austin, 1978; Sax, Gilmartin, and Bryant, 2003), education (Billiet et al., 2007; Sax, Gilmartin, and Bryant, 2003), and minority status (Feskens et al., 2006; Feskens et al., 2007). Against the background of declining response rates, researchers employ various strategies to encourage survey participation, namely mixed-mode surveys, increasing the number of contacts with invited participants, and unconditional (monetary) incentives (Becker, Möser, and Glauser, 2019; Dillman, 2007; Groves and Peytcheva, 2008). Nevertheless, the selective response behavior and its implication for the generalization of results to the population of interest remains a core concern for survey-based research.

At the same time, the characteristics that determine people's response in social surveys are also associated with their place of residence. Facing different opportunities and constraints (e.g., the individually available monetary resources or housing alternatives on the market), people actively choose where to live, leading to more or less stable levels of social and ethnic residential segregation (e.g., De la Roca, Ellen, and O'Regan, 2014; Musterd, 2005; Musterd et al., 2017; Zangger, 2019). Consequently, people's response behavior in social surveys is spatially correlated. In other words, a particular unit's response to an invitation to participate in a survey is similar to the one of other people in its vicinity who were also invited. This spatial dependency has been recognized and researchers use, for example, geographic information systems (GIS) to map unit nonresponse in surveys and to identify areas with especially low participation rates in social surveys (Abbott and Compton, 2014; N. Bates and Mulry, 2011; Hansen et al., 2007). Related to this insight in the geographic correlates of survey nonresponse, researchers have turned to widely available aggregated measures to construct survey weights to minimize the impact of nonresponse bias, either by means of including aggregated measures and random effects when calculating inverse-probability weights or by means of calibration (Erdman and N. Bates, 2017; Kreuter, Olson, et al., 2010; Skinner and D'arrigo, 2011; Valliant, Dever, and Kreuter, 2018). In this regard, it has been shown that including aggregated measures at (preferably) small geographic scales can indeed decrease nonresponse bias—at least

if these characteristics are sufficiently associated with both the response propensity and the survey variables of interest (Biemer and Peytchev, 2013; Kreuter, Olson, et al., 2010; Olson, 2013).

This paper takes the spatial correlation of (non-)response as a starting point and investigates how including information about neighboring units who were also invited might be a more direct approach to improve survey weights and to reduce the bias from survey nonresponse. While existing research using response-propensity weighting largely relies on random effect and related models to account for the spatial clustering of nonresponse (Skinner and D’arrigo, 2011), this study proposes and examines spatial econometric models to account for spatial autocorrelation in people’s response behavior. Spatial models directly model the mutual spatial dependence among units at the individual level rather than using aggregated (census) characteristics of predefined administrative areas. Doing so, they offer a means for integrating spatial dependence into the estimation of response propensities without the need to fully observe the spatial selection process. This is achieved by allowing the response propensity of a particular unit i to depend on the response of other units j in i ’s vicinity, either by means of a residual spatial correlation in the error term or through lags of the dependent (response) or independent model variables (Cressie, 2015; LeSage and Pace, 2009). Using Monte Carlo simulations, this paper examines how such models are an effective alternative for constructing propensity weights to adjust for selective non-response when the underlying response behavior of invited participants is unknown. Moreover, we compare the spatial approach to alternative strategies, namely random effect and ordinary logistic regression models.

The remainder of this paper is organized as follows. The next section presents the background and rationale of using spatial models and reviews existing research. The third section elaborates on the simulation setup and the different model specifications. The results of the simulations are presented in the fourth section. The paper concludes with a discussion of the use of spatial models to accommodate for unit nonresponse in social surveys.

2. Background

To deal with unit nonresponse, researchers have turned to various techniques. Broadly speaking, two strategies are frequently employed: nonresponse propensity weighting (NPW) and nonresponse calibration (NC) (Brick, 2013; Haziza and Lesage, 2016; C.-E. Särndal and Lundström, 2005; E. Särndal, 2007; Valliant, Dever, and Kreuter, 2018). NPW uses observed characteristics of both respondents as well as nonrespondents to predict the individual response probability, ϕ_i . On the other hand, calibration makes use of known population totals to adjust respondents' marginal distribution to allow for a generalization to the population of interest (e.g., poststratification or raking). Both approaches aim at reducing nonresponse bias, consistent and efficient estimates, and can—ideally—be combined when propensity weights are used as the starting point for calibration, resulting in a two-step procedure (Haziza and Lesage, 2016; Laaksonen, 2018).

In this contribution, we make some simplifying assumption. First of all, we only look at nonresponse propensity weighting. Second, we assume that the data are obtained by simple random sampling. While the latter makes things easier, it also might not adequately represent the strategies used in many surveys. Nevertheless, the underlying reasoning of using spatial information to predict survey nonresponse also applies to more complex sampling schemes and is thus of value beyond the following illustration assuming simple random sampling.

The idea of using geographic information in the estimation of response probabilities is not new. On the one hand, only few characteristics are usually known of both respondents and nonrespondents. This makes it necessary to use aggregated auxiliary information (Kreuter, Couper, and Lyberg, 2010; Kreuter, Olson, et al., 2010; Olson, 2013). Such auxiliary information, for example, regarding obtained educational degrees, is often available at different geographic entities—from census tracts to districts or zip-code areas. The aggregated characteristics are then entered as predictors to obtain individual response propensities. On the other hand, it has more generally been noted that nonresponse varies geographically and that mapping the patterns of nonresponse are a means to better understand the underlying process (N. Bates

and Mulry, 2011; Erdman and N. Bates, 2017). Biemer and Peytchev (2013) go one step further and demonstrate that adding geocoded census data at the block-group or tract level contributes little in reducing nonresponse bias unless the characteristics are highly geographically clustered and exhibit a good correlation with the outcome measures of interest.

In this contribution, we take the observation of a spatial clustering and correlation of nonresponse as a starting point. As mentioned above, we look at the case of simple random sampling from a known population. In this case, base weights are just the inverse of the inclusion probability $\pi_i = n/N$, where N is the number of units in the population and n the number of sample units, $i \in 1, \dots, n$. Since base weights are constant across the units in the sample, we omit them from the following presentation. Instead, we focus on the propensity score, that is, the probability of responding to being invited to the survey (Rosenbaum and Rubin, 1983). The response propensity for a unit $i \in 1, \dots, n$ is given as

$$\phi_i = \phi_i(\mathbf{x}_i) = Pr(R_i = 1 | \mathbf{X} = \mathbf{x}_i), \quad (1)$$

where $R_i = 1$ indicates a response and $\mathbf{x}_i$ is a vector of observed characteristics for all units in the sample (both respondents as well as nonrespondents—Brick, 2013). There is much research on the auxiliary variables in $\mathbf{X}$ that can or should be used to predict survey nonresponse (see, for example, Bethlehem, Cobben, and Schouten, 2011; Groves and Couper, 1998; Kreuter, Couper, and Lyberg, 2010; C.-E. Särndal and Lundström, 2008; Schouten, 2007). Given the response propensity in (1), an estimator of the population mean of Y , for example, is then given as

$$\hat{y}_r = \frac{1}{N} \sum_{i \in S_r} d_i \phi_i^{-1} y_i. \quad (2)$$

S_r denotes the set of respondents in the sample and d_i are the design weights briefly introduced before. Common choices to estimate ϕ_i , the individual response propensity which are often grouped into response classes (Brick, 2013; Valliant, Dever, and Kreuter, 2018), include logit and probit models. These models assume that observations are i.i.d., that is, identically and independently distributed. Spatial models

relax this assumption and are briefly introduced in the next section before we move on to evaluating how they can be used to incorporate the spatial dimension of the nonresponse mechanism.

2.1. Spatial models

Spatial models differ from traditional regression models in that they allow observations to be interdependent. That is, these model do not assume that observations are (conditionally) independent but they are allowed to mutually influence each other (Kelejian and Piras, 2017; LeSage and Pace, 2009). They are thus especially suited to model social processes such as social interactions, networks, or diffusion processes (Beck, Gleditsch, and Beardsley, 2006).

The interdependence among observation is reflected in the spatial weights matrix $\mathbf{W}$. Its elements $w_{ij} \in \mathbf{W}$, $i = 1, \dots, n$, reflect the extent to which two units i and j , $i \neq j$, are mutually dependent. They can either represent binary information whether the two elements are mutually dependent (adjacency matrix) by setting

$$w_{ij} = \begin{cases} 1, & \text{if interdependent} \\ 0, & \text{otherwise,} \end{cases}$$

or more sophisticated information about the extent and direction of the dependence by allowing $w_{ij} \in \mathbb{R}$. A common choice is the inverse distance matrix in which case the elements of $\mathbf{W}$ weight the mutual influence of two observations by their “closeness”: the farther apart, the less they influence each other. This latter approach is also taken in the following simulation.

In these models, the spatial dependence can work through three different processes. First, there might be spatial lags of the outcome measure (the term $\mathbf{WY}$ in equation (3)). In this case, the outcome of one unit is not only determined by its independent characteristics but also by the outcome of neighboring units which in turn are also influenced by the outcome of that particular unit. The scalar ρ depicts the corresponding coefficient. In the terminology of Manski (1993), this would refer to an

endogenous contextual effect.

$$\begin{aligned}
\mathbf{Y} &= \alpha\mathbf{I} + \rho\mathbf{W}\mathbf{Y} + \mathbf{X}_{\mathcal{A}}\beta + \mathbf{W}\mathbf{X}_{\mathcal{B}}\theta + \boldsymbol{\nu} \\
\boldsymbol{\nu} &= \lambda\mathbf{W}\boldsymbol{\nu} + \boldsymbol{\varepsilon} \\
\boldsymbol{\varepsilon} &\sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})
\end{aligned} \tag{3}$$

Second, the term $\mathbf{W}\mathbf{X}_{\mathcal{B}}$ depicts exogenous contextual effects in Manski’s (1993) terminology. Some characteristics of respondent j might directly influence the outcome of a unit i and vice versa. The subscript $\mathcal{B}$ reflects that these variables might be other characteristics than the “purely” individual ones, $\mathbf{X}_{\mathcal{A}}$. Usually, $\mathbf{X}_{\mathcal{B}} \subset \mathbf{X}_{\mathcal{A}}$. Finally, there might be a spatial correlation of the errors, depicted by the λ coefficient. Such a spatial correlation, which can be understood as correlated effects according to Manski (1993), can result from the unobserved selection of people into different contexts or the omission of important, spatially clustered, individual predictors in $\mathbf{X}_{\mathcal{A}}$ (LeSage and Pace, 2009; Vega and Elhorst, 2015).

2.2. Spatial models for predicting nonresponse

The basic idea of this contribution is quite simple: Given the often documented spatial clustering of survey nonresponse (e.g., N. Bates and Mulry, 2011; Biemer and Peytchev, 2013; Erdman and N. Bates, 2017; Hansen et al., 2007), we might wonder to which extent using the information about other units participation status, R_j , in the vicinity of an observation i might improve the prediction of unit i ’s participation status, R_i . Thus, we look for a model of the form

$$\phi_i = \Pr(R_i = 1 | \mathbf{x}_i, R_j) = f(\mathbf{x}_i, R_j). \tag{4}$$

R_i is an individual’s response status. It equals 1, when they responded to the survey and 0 otherwise. $\mathbf{x}_i$ is a vector of observed characteristics for both respondents and nonrespondents.

Following a random response model (Bethlehem, Cobben, and Schouten, 2011; Brick, 2013), a natural choice to incorporate the information about other units’ re-

sponse status would thus be a spatial logit or probit model (Franzese, Hays, and Cook, 2016). Using latent variable representation with $R_i = 1$ if $R_i^* > 0$, such a model takes the form

$$\begin{aligned} \mathbf{R}^* &= \rho \mathbf{W} \mathbf{Y}^* + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}, \end{aligned} \tag{5}$$

where the second line in equation (5) gives the reduced form and $(\mathbf{I} - \rho \mathbf{W})^{-1}$ is called the spatial multiplier. Assuming a probit model, the probability for a response is given as

$$\begin{aligned} \phi_i(\mathbf{x}_i) &= p\left([\mathbf{X} \boldsymbol{\beta}]_i + [(\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}]_i > 0\right) \\ &= \Phi\left\{[\mathbf{X} \boldsymbol{\beta}]_i / \sigma_{\varepsilon_i}\right\}, \end{aligned} \tag{6}$$

where $\Phi(\cdot)$ is the cumulative normal distribution (Franzese, Hays, and Cook, 2016).

However, does such a model adequately represent the underlying missing mechanism? From the observed clustering in empirical applications, we might conclude that it indeed does so. However, nonresponse is likely not a “contagious” behavior as suggested by the data generating process (DGP) in equation (5). Rather than a true effect of others’ participation status, unobserved characteristics that also influence one’s place of residence are a more likely explanation for the geographic clustering of nonresponse in the empirical world.

If, however, there are some unobserved characteristics, $\mathbf{Z}$, that influence both participation status and geographic location, a more adequate model would be one with spatially correlated errors rather than a spatial lag. Using again the latent variable formulation as before, the reduced form of a spatial error model takes the form

$$\mathbf{R}^* = \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}. \tag{7}$$

However, since $\mathbb{E}[\boldsymbol{\varepsilon}] = 0$, the spatial information does not enter in the predicted probabilities $\hat{\phi}_i$. Thus, although it is the adequate model from an omitted variable perspective, it does not help us in incorporating the information about others’ participation status in a substantial way. This is why we are going to focus on the potential

of spatial lag models to improve response predictions, that is, the model in equation (6).¹ We will also look at the adequacy of a spatial lag model using robust lagrange multiplier tests (Anselin et al., 1996).

3. The simulation

Before we proceed to the results, we need to briefly outline the simulation approach taken in this study. As a simplifying assumption, we only use two observed and one unobserved variable which influence people’s response probability. Moreover, we assume that the marginal distribution of the unobserved distribution is known in specific geographical areas, as it is often the case in real world scenarios (e.g., we likely do not know the educational certificates people obtained prior to the survey but we often have information about the, for example, the share of people with a university degree in prespecified census blocks).

3.1. The setup

We assume that there are three covariates that—to the degree of some random noise—determine the likelihood people will participate in a survey. Two of them are themselves correlated² and stem from a multivariate normal distribution of the form

$\mathbf{X} \sim \mathcal{N}_2(\mathbf{1}, \Sigma)$, where

$$\Sigma = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}.$$

---

¹Alternatively, we could also look at spatial Durbin or SLX models, that is, models that only include a lag of independent variables— $\mathbf{W}\mathbf{X}_B\boldsymbol{\theta}$ (Vega and Elhorst, 2015). These models could be a worthwhile alternative as long as there are observed characteristics of both respondents and nonrespondents that are associated with the response probability, the outcome measure of interest, and the unobserved variables $\mathbf{Z}$ that also affect geographic location.

²In the case presented here, the two variables are positively correlated. Figures A.1 and A.2 in the appendix show comparable results when the two variables are negatively correlated with

$$\Sigma = \begin{bmatrix} 1 & -0.2 \\ -0.2 & 1 \end{bmatrix}.$$

Figure 1: Spatial variation of aggregated $u\bar{X}_1$

The two normally distributed variables are constructed to be spatially correlated using the spatial multiplier defined above. To this end, the two variables are replaced by

$$\begin{aligned} \mathbf{X}^* &= (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}, \\ \rho &= \{0.4, 0.5, 0.6\} \end{aligned} \tag{8}$$

and are then transformed to a Binomial and a Bernoulli distributed variable via their distribution function. The resulting two variables are

$$\begin{aligned} uX_1 &\sim \text{Bin}(5, 0.2) \\ oX_1 &\sim \text{Ber}(0.01), \end{aligned}$$

where uX_1 is the unobserved, binomial-distributed variable and oX_1 is the observed Bernoulli variable. We might think of the unobserved variable uX_1 as people's education and the observed one, oX_1 , as an indicator for a migration background which often is known from population registers for both respondents as well as nonrespondents. Moreover, we assume that there is an additional observed variable, oX_2 , which follows a Binomial distribution of the form $oX_2 \sim \text{Bin}(3, 0.4)$.

In the simulation, we randomly place 1'000 observations on a 50 times 50 grid. The three variables above are used for the data generating process, which takes the form

$$\phi_i = \frac{\exp(\beta_0 + \beta_{uX_1} \times uX_1 + \beta_{oX_1} \times oX_1 + \beta_2 \times oX_2)}{(1 + \exp(\beta_0 + \beta_{uX_1} \times uX_1 + \beta_{oX_1} \times oX_1 + \beta_2 \times oX_2))}, \tag{9}$$

where

$$\beta_0 = \{-1.8, -2.0, -2.2\}$$

$$\beta_{uX_1} = 1.7$$

$$\beta_{oX_1} = -2.5$$

$$\beta_{oX_2} = -2.5.$$

These values result in an average response propensity of about 44% to 45%, ranging from 0.01% to 99.82%. Indicated by the three values of ρ in equation (8), the extent of the underlying spatial correlation will also be varied. To keep the other coefficients constant in the simulations, we needed to adjust the intercept accordingly such that each of the three levels of spatial correlation still result in about the same distribution of the response propensity.

The 50×50 grid is divided into 25 and 100 subgrids, respectively. For each of these 25 and 100 level 2 units, the average of uX_1 , $\bar{uX}_1$, is computed as an additional covariate. This is intended to reflect that in real world scenarios, we often have aggregated information on some unobserved characteristics of the people in our sample which can be used to estimate response propensities. The geographical distribution of this aggregated measure over the subgrids is depicted in Figure 1.

To evaluate the efficiency and consistency of weights derived from spatial models, we focus on the following outcome variable:

$$\begin{aligned} y_i &\sim N(\mu_y, \sigma_y), \\ \mu_y &= 3000 + 300 \times uX_1 - 950 \times oX_1 + 200 \times oX_2 \\ \sigma_y &= 300 \end{aligned} \tag{10}$$

If we think of uX_1 as people's education and oX_1 as an indicator of a migration background, y_i can be taken, for example, as their income, and oX_2 something like their labor market experience. To keep things simple, we will focus on the weighted income, using the formula in equation (2).

3.2. Analytical approach

The setup described above is repeated 500 times, the number of iterations in the simulation. In each step, we are going to use the predicted probabilities from three different types of models. First, we estimate ordinary logit models with only oX_1 and oX_2 as predictors. Second, spatial models with a) a spatial lag of the dependent variable ($\rho\mathbf{WY}$) and b) with spatially correlated errors ($\lambda\mathbf{W}\varepsilon$) are estimated. Third, we add the average of uX_1 at the subgrid level as an additional predictor and estimate ordinary random effect models (generalized linear and mixed models—Pinheiro and D. Bates, 2006) along with the two spatial models from the second step. Since we evaluate 6 different scenarios (3 levels of ρ and two different subgrids within each) and use 6 different models, we estimate a total of $6 \times 6 \times 500 = 18'000$ models in $6 \times 500 = 3'000$ simulations.

To evaluate the performance of the different estimators, we look at two different measures. On the one hand, we are going to judge their adequacy by the average share of correctly classified observations. On the other hand, we use the mean squared error (MSE) when comparing the true average outcome $\bar{Y}$ to the one obtained using propensity weighting. MSE is a measure that takes into account both variance (efficiency) as well as bias (Carsey and Harden, 2014). It is given as

$$MSE_y = \mathbb{E} \left[\left(\hat{Y} - \bar{y} \right)^2 \right] = \mathbb{E} \left[\left(\hat{y} - \mathbb{E} [\hat{y}] \right)^2 \right] + \left(\mathbb{E} [\hat{y}] - \bar{Y} \right)^2, \quad (11)$$

where the last two terms reflect variance and bias, respectively. We will mainly rely on a graphical representation of the results from all the simulations. The code for the simulations can be found in the appendix.

4. Results

We start by an examination of the extent to which the simulated data does indeed depict the intended spatial correlation. We constructed the two correlated measures uX_1 and oX_1 to also be spatially correlated, reflecting the fact that similar people tend to live in similar places. Since the response propensity, ϕ_i , is partially deter-

Figure 2: Spatial correlation in units' response status, R_i , and the outcome measure, y_i

mined by these two variables, units' response status, R_i , should also be spatially correlated, although this correlation is entirely due to the (unobserved) independent variables. The two variables, which we can think of as people's education and migration background, for example, also enter the DGP of the outcome measure, y_i , thought of as people's income, which should thus result in a spatial correlation in y_i among neighboring units.

We test whether units' dichotomous response status, $R_i = 1$, and the outcome measure, y_i , are indeed spatially correlated in the simulated data using Moran's I . Moran's I is the spatial equivalent to the correlation coefficient (Cressie, 2015) and is defined as

$$I = \frac{n}{\sum_i \sum_j w_{i,j}} \frac{\sum_i \sum_j w_{i,j} x_i x_j}{\sum_i x_i^2}, \quad (12)$$

where $w_{i,j}$ is the spatial weight for the association between two units i and j , $i \neq j$. Figures 2a and 2b depict the distribution of the Moran's I test statistic across all the simulations. The dash-dotted lines in Figure 2 represents the expected value of the Moran's I test statistic under the null-hypothesis of no spatial association among observations. Since the empirical Moran's I test statistics are located to the right of this value in more than 95% of all the simulations, we conclude that there is a positive spatial correlation in both the response indicator, R_i , as well as the outcome, y_i . Given the dichotomous nature of R_i , the somewhat lower spatial correlation comes as no surprise. Likewise, we witness a stronger spatial correlation with a more pronounced spatial dependence in the DGP (ρ coefficient; comparing the different panels of Figures 2a and 2b).

Figure 3: Model performance: Share correctly classified

In a first step to evaluate the extent to which spatial models can be used to obtain adequate propensity weights, we look at how well the different estimators predict a unit's response status R_i . To this end, we look at 6 different models. The first three only use the observed unit-level covariates oX_1 and oX_2 as predictors. The estimators are: a GLM model for binary dependent variables (ordinary logistic regression model), a spatial error, and a spatial lag model, both for binary dependent variables (spatial probit models). The other set of estimators also include the average of the unobserved covariate, $u\bar{X}_1$, as an additional contextual regressor in either 25 or 100 subgrids (level 2 units). These three models comprise a generalized linear mixed model (GLMM), a spatial error, and a spatial lag model, respectively.

From Figure 3, depicting the results from 500 iterations in the simulation, we first note that including the average of the unobserved covariate uX_1 , that is, $u\bar{X}_1$, improves the fit to the data considerably, even though this measure is not part of the DGP of R_i . While the increase is no more than 1.5 percentage point when the artificial world is divided into 25 different level 2 units, the share of units for whom we correctly predict the observed response status, R_i , increases by about 3 percentage points if the aggregated information, $u\bar{X}_1$, is calculated at a finer-grained level of 100 subgrids. This comes as no surprise since with a geographical resolution, that is, smaller subgrids, we move closer to capturing the underlying spatial correlation by means of including $u\bar{X}_1$ in the estimation. This, however, seems to hold true

irrespective of the strength of the underlying spatial association in the data generating process: The improvement in correctly predicting units' response status when including $u\bar{X}_1$ in different level 2 units is mostly independent of ρ .

As expected, Figure 3 also reveals no improvement in predicting people's response status when using a spatial error model (SEM) instead of an ordinary logistic regression model (GLM). Although a spatial error model should more adequately reflect the data generating process, given that R_i is spatially dependent due to the omission of uX_1 from the regression equation (LeSage and Pace, 2009; Vega and Elhorst, 2015), it cannot alter much the predictions since the spatial association only enters the error term, which by definition is omitted from the predictions. On the other hand, the spatial lag model seems to fit the data somewhat worse. When we use neighboring units' observed participation status (from the 5 closest neighbors, using inverse distance weights, w_{ij}), the share of correctly classified observations is slightly lower. Although a misspecification of the spatial process (spatial lag instead of a spatial error) can generally explain the worse performance of the SAR estimator, a closer examination reveals that this is due to the setup of the simulation: Since the response rate was set to slightly above 40%, the majority of cases are nonrespondents in the simulation. Given the positive spatial correlation in the DGP (ρ values of 0.4, 0.5, and 0.6, respectively), nearby units j tend, on average, to depress a unit i 's response propensity. There are only negligible differences among the three estimators when we additionally add the contextual characteristic $u\bar{X}_1$ to predict units' response status. In this case, using a spatial lag model seems to be a slightly better choice than a spatial error model, although it is no better than just using a mixed effect (multilevel) model.

Based on model performance in terms of correctly predicting units' response status, we would thus expect spatial models to derive no better weights to adjust for nonresponse than ordinary logistic or random effects models. Before we move on, however, we briefly examine whether there is some spatial correlation left after controlling for observed factors in the non-spatial models. If they are, this might bias coefficients and thus the predicted response probability (Pace and LeSage, 2010). Since the GLM model does not account for the spatial dependence other than through the one in-

Figure 4: Residual spatial correlation

duced by oX_1 , it comes as no surprise that the residuals of such a regression are still spatially dependent (Figure 4a). However, even if we capture the spatial process at an aggregated level by means of including $u\bar{X}_1$, the average of the unobserved, spatially correlated covariate at level 2, we still observe considerable spatial dependence in the residuals of the GLMM models (Figure 4b). This is especially the case if the geographical resolution is lower, that is, the contextual information is available on a more aggregated level of 25 instead of 100 subgrids. Consequently, a spatial lag or error model which also incorporates the contextual regressor $u\bar{X}_1$ might be more adequate to obtain propensity weights since they will account for the residual spatial correlation.

Figure 5 depicts the results from the simulations when we use the inverse of the estimated response propensity, that is, $1/\hat{\phi}_i$, to weight the outcome measure y_i to obtain an estimate of the average outcome in the population, $\hat{\bar{Y}}$. To evaluate the extent to which the weighting is successful, we look at the mean squared error (MSE), reflecting the deviation from the true value $\bar{Y}$. Thus, we look for the estimator which yields the lowest average MSE across simulations. Without any further adjustment of the inverse probability weights, the estimators' performance is quite heterogeneous. While inverse probability weights obtained by an ordinary logistic regression model (GLM in Figure 5) seem to be a good choice when the underlying spatial association is not too strong ($\rho = 0.4$), these weights perform quite badly when the underlying spatial correlation is more pronounced ($\rho = 0.6$). There seems to be no unique pattern apart from the fact that the unadjusted inverse probability weights obtained through a

Figure 5: Mean squared error of the estimated average outcome, propensity weights

spatial error or spatial lag model which also include the aggregated measure $\bar{uX}_1$ seem to yield comparably lower MSE values, indicating less bias and/or a smaller variance. Since the residuals of an ordinary multilevel (mixed effects) model (GLMM) which includes the average of the unobserved covariate uX_1 , that is, $\bar{uX}_1$, are still spatially correlated (Figure 4b), it comes as no surprise that the average MSE is usually higher than the one from spatial lag and spatial error models which also include $\bar{uX}_1$ as an independent regressor.

As in many real-world scenarios, the regression models are misspecified since they do not include the unobserved covariate uX_1 , which is part of the data generating process. In this regard, the literature highlights the importance of weight trimming (Haziza and Beaumont, 2017; Lee, Lessler, and Stuart, 2011): If estimators perform poorly with respect to the data generating process, propensity weights will be quite disperse, increasing the variance of the estimates of interests. However, the smaller variance through weight trimming is usually obtained at the expense of introducing (additional) bias (Haziza and Beaumont, 2017). Nevertheless, trimming large weights can improve the performance of propensity scores, although the increase in efficiency depends on the estimator used (Lee, Lessler, and Stuart, 2011). To evaluate the

Figure 6: Mean squared error of the estimated average outcome, trimmed weights

impact of weight trimming, the maximum weight was set to 20, which is equivalent to setting the cutpoint at the 90th percentile, reflecting moderate weight trimming. The original propensity weights yielded weights over 200.

Figure 6 depicts the MSE of the estimated average outcome using the trimmed weights. As expected, the trimming generally decreases variance, indicated by the smaller confidence intervals. Given the tradeoff of introducing bias through the weight trimming, it comes as no surprise that the average MSE did not really decrease when using the trimmed weights (comparing Figures 5 and 6). This indicates that the trimming might be too conservative with regard to large weights. Nevertheless, the picture of different estimators' performance becomes much clearer once extreme weights are trimmed. First of all, the more pronounced the spatial correlation among units (increasing values of ρ), the better the performance of any of the estimators. This is due to the fact that the observed covariate oX_1 is correlated with the unobserved uX_1 , picking up shares of the underlying spatial process. When additional information is available at the aggregated level, such as the one obtained from census or administrative data for census blocks or wards, spatial models including this information (i.e., $u\bar{X}_1$) generally outperform ordinary multilevel models (GLMM). This is especially true when there is a smaller number of level 2 (aggregated) units. Given that there are only 1000 observations that are randomly distributed over the 50×50 grid, a number of 25 subgrids (level 2 units) is more likely to match the data available in

real life applications. In such cases, using spatial models with trimmed propensity weights yields considerably better estimates of weighted population averages. These models pick up the remaining spatial correlation after including level 2 characteristics ($u\bar{X}_1$). Although spatial error and spatial lag models yield about the same average MSE in these scenarios, a spatial error model more adequately represents the underlying data generating mechanisms since they model the spatial process as one that is due to unobserved, spatially correlated covariates rather than suggesting a “contagion mechanism” as in the case of spatial lag models. This is then also represented in the slightly lower variance, indicated by smaller confidence intervals for the average MSE in Figure 6.

5. Discussion

This contribution examined the potential of spatial econometric models to obtain propensity weights for unit-nonresponse in social surveys (LeSage and Pace, 2009). Starting from the observation that patterns of nonresponse are often geographically correlated (N. Bates and Mulry, 2011; Biemer and Peytchev, 2013; Erdman and N. Bates, 2017), a Monte Carlo simulation with a spatially correlated response pattern was set up. In the simulation, the response probability was defined to be influenced by two spatially correlated covariates, one observed and one unobserved (oX_1 and uX_1 , respectively). These two covariates were themselves correlated such that the observed variable picks up some of the spatial correlation of the unobserved one. Moreover, an additional, spatially uncorrelated covariate, oX_2 , also influenced the response propensity, $\phi_i(\mathbf{X})$. The same three measures were also set to influence the outcome measure, y_i .

To examine the behavior of propensity weights obtained from spatial model, we looked at the mean squared error (MSE) of the estimated population average of the outcome, that is, $\hat{\bar{Y}}$. To this end, the weighted estimates obtained from spatial error and spatial lag models were compared to the ones from non-spatial models. The latter included ordinary logistic regression models as well as mixed effect models in which $u\bar{X}_{1j}$, the average value of the unobserved covariate in level 2 unit j , was

included as an additional regressor. Doing so reflects the fact that aggregated population characteristics are often among the auxiliary information available to estimate response propensities in real world applications (Kreuter, Olson, et al., 2010; Olson, 2013). In the simulations, the strength of the underlying spatial process as well as the geographic resolution of the level 2 units was artificially altered.

From a theoretical point of view, the omission of a spatially correlated covariate should best be picked up by a spatial error model (LeSage and Pace, 2009; Vega and Elhorst, 2015). However, since the spatial process works entirely through the residuals of such models, predictions from spatial error models differ from ordinary regression models only to the extent as the observed part of the model (the estimated coefficients) is affected by the omission of the spatial process. It was thus argued that a spatial lag model, that is, a model in which the dependent variable of a unit i is affected by the dependent variable of neighboring units j , treating units' response behavior as if it were contagious, might be a more efficient way to include the spatial dependency into the predictions.

Results indicate that – as expected – spatial error models perform no better than ordinary regression models when it comes to predicting unit nonresponse, irrespective whether the models include the aggregated level 2 information about the unobserved level 1 covariate. Contrary to the expectation, however, spatial lag models generally perform worse when it comes to correctly predicting a unit's response status. This can, however, be attributed to the setup of the simulation: The response rate was set to be slightly higher than 40% across simulations. Thus, on average, units will be nonrespondents and given the positive spatial correlation among units, a particular unit i 's response propensity will, on average, be depressed by neighboring units response status. In other scenarios where the majority of units respond to the survey, spatial lag models might thus outperform other type of models—if survey response is spatially correlated.

When we move on to estimating the weighted average outcome in the simulations, a somewhat different picture emerges. Especially if weights are trimmed to take into account the large variance that results from a misspecification inherent to all the models (omitted variable bias), the weights from spatial models generally yield better

weights than using the inverse of the estimated response propensity from ordinary logistic and nonlinear mixed effects models. While the differences are small if there are many level 2 units, that is, if the aggregated information about the unobserved covariate is available at a small geographical scale, they are quite pronounced if there are only a few level 2 units and the aggregated information is thus less precise. In this case, spatial error and spatial lag models both yield a considerably lower MSE across simulations, irrespective of the strength of the underlying spatial association. However, given their better fit to the data generating process, spatial error models generally yield a somewhat smaller variance.

Spatial models can thus indeed yield better propensity weights to correct for unit nonresponse. However, their ability to do so depends on several assumptions: Most importantly, survey response needs to be spatially correlated. While this can generally be expected to be the case in the real world since people are selected into different residential areas based on characteristics that are often not known prior to data collection, this nevertheless does not always need to be the case. Moreover, all the simulations assumed a moderate spatial association in the unobserved covariate, ranging from $\rho = 0.4$ to $\rho = 0.6$. Likewise, a rather low response rate of roughly 40% was assumed. Although similar response rates can be encountered in many contexts, results might nevertheless also depend on this additional assumption. The better performance of the spatial models is further only evident if aggregated information at a higher level such as census tracts or wards is available and especially if these level 2 units have a lower spatial resolution, that is, they are geographically bigger and thus more heterogeneous. Consequently, similar to approaches that use aggregated, geocoded census data (Biemer and Peytchev, 2013), propensity weights obtained by means of spatial models can improve nonresponse adjustments, although the improvement to established methods is rather moderate.

Given their dependence on the assumptions outlined above as well as additional ones (namely, the specification of the underlying weights matrix—LeSage and Pace, 2014), future research should carefully evaluate these assumptions to test whether spatial models indeed keep their theoretically assumed promises. Likewise, their ability for robust inference should be examined and compared to existing approaches using real-

world data. Finally, the Monte Carlo simulations were based on some simplifying assumptions (e.g., simple random sampling). To assess the true potential of spatial models for nonresponse adjustments, future research might thus want to examine their behavior in different settings and when combining them with additional nonresponse adjustment techniques, namely, calibration or weighting class adjustments (Valliant, Dever, and Kreuter, 2018).

References

- Abbott, Owen and Garnett Compton (2014). “Counting and estimating hard-to-survey populations in the 2011 census”. In: *Hard-to-Survey Populations*. Ed. by Roger Tourangeau et al. Cambridge: Cambridge University Press, pp. 58–81.
- Abraham, Katharine G, Aaron Maitland, and Suzanne M Bianchi (2006). “Nonresponse in the American Time Use Survey”. In: *Public Opinion Quarterly* 70.5, pp. 676–703.
- Anselin, Luc et al. (1996). “Simple diagnostic tests for spatial dependence”. In: *Regional science and urban economics* 26.1, pp. 77–104.
- Bates, Nancy and Mary H Mulry (2011). “Using a Geographic Segmentation to Understand, Predict, and Plan for Census and Survey Mail Nonresponse”. In: *Journal of Official Statistics* 27.4, pp. 601–618.
- Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley (2006). “Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy”. In: *International Studies Quarterly* 50.1, pp. 27–44.
- Becker, Rolf, Sara Möser, and David Glauser (2019). “Cash vs. vouchers vs. gifts in web surveys of a mature panel study—Main effects in a long-term incentives experiment across three panel waves”. In: *Social Science Research* Online First.
- Bethlehem, Jelke G., Fannie Cobben, and Barry Schouten (2011). *Handbook of nonresponse in household surveys*. Wiley Series in Survey Methodology. OCLC: ocn611036930. Hoboken, N.J: John Wiley & Sons.
- Biemer, P. and A. Peytchev (2013). “Using Geocoded Census Data for Nonresponse Bias Correction: An Assessment”. In: *Journal of Survey Statistics and Methodology*

- 1.1, pp. 24–44. URL: <https://academic.oup.com/jssam/article-lookup/doi/10.1093/jssam/smt003>.
- Billiet, Jaak et al. (2007). “Estimation of Nonresponse Bias in the European Social Survey: Using Information from Reluctant Respondents”. In: *Journal of Official Statistics* 23.2, pp. 135–162.
- Brick, J. Michael (2013). “Unit Nonresponse and Weighting Adjustments: A Critical Review”. In: *Journal of Official Statistics* 29.3, pp. 329–353.
- Carsey, Thomas M. and Jeffrey J. Harden (2014). *Monte Carlo Simulation and Resampling Methods for Social Science*. Thousand Oaks: Sage.
- Cressie, Noel A. (2015). *Statistics for Spatial Data*. New York: Wiley.
- Criqui, Michael H., Elizabeth Barrett-Connor, and Melissa Austin (1978). “Differences between respondents and non-respondents in a population-based cardiovascular disease study”. In: *American Journal of Epidemiology* 108.5, pp. 367–372.
- De la Roca, Jorge, Ingrid Gould Ellen, and Katherine M. O’Regan (July 2014). “Race and neighborhoods in the 21st century: What does segregation mean today?” In: *Regional Science and Urban Economics* 47, pp. 138–151.
- Dillman, Don A. (2007). *Mail and Internet Surveys. The Tailored Design Method*. 2nd ed. Hoboken, NJ: John Wiley & Sons.
- Erdman, Chandra and Nancy Bates (2017). “The Low Response Score (LRS)”. In: *Public Opinion Quarterly* 81.1, pp. 144–156.
- Etter, Jean-François and Thomas V. Perneger (1997). “Analysis of non-response bias in a mailed health survey”. In: *Journal of Clinical Epidemiology* 50.10, pp. 1123–1128.
- Feskens, Remco et al. (2006). “Collecting Data among Ethnic Minorities in an International Perspective”. In: *Field Methods* 18.3, pp. 284–304.
- (2007). “Nonresponse Among Ethnic Minorities: A Multivariate Analysis”. In: *Journal of Official Statistics* 23.3, pp. 387–408.
- Franzese, Robert J., Jude C. Hays, and Scott J. Cook (2016). “Spatial- and Spatiotemporal-Autoregressive Probit Models of Interdependent Binary Outcomes”. In: *Political Science Research and Methods* 4.1, pp. 151–173.

- Groves, Robert M. and Mick Couper (1998). *Nonresponse in household interview surveys*. Wiley series in probability and statistics. New York: Wiley.
- Groves, Robert M. and Emilia Peytcheva (2008). “The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis”. In: *Public Opinion Quarterly* 72.2, pp. 167–189.
- Hansen, Richard A. et al. (2007). “Geographic Information System mapping as a tool to assess nonresponse bias in survey research”. In: *Research in Social and Administrative Pharmacy* 3.3, pp. 249–264.
- Haziza, David and Jean-François Beaumont (2017). “Construction of Weights in Surveys: A Review”. In: *Statistical Science* 32.2, pp. 206–226. URL: <http://projecteuclid.org/euclid.ss/1494489812> (visited on 08/28/2020).
- Haziza, David and Éric Lesage (2016). “A Discussion of Weighting Procedures for Unit Nonresponse”. In: *Journal of Official Statistics* 32.1, pp. 129–145.
- Kelejian, Harry H. and Gianfranco Piras (2017). *Spatial Econometrics*. London: Academic Press.
- Kreuter, Frauke, Mick Couper, and Lars Lyberg (2010). “The Use of Paradata to Monitor and Manage Survey Data Collection”. In: *Proceedings of the Survey Research Methods Section of the American Statistical Association*, pp. 282–296.
- Kreuter, Frauke, Kristen Olson, et al. (2010). “Using proxy measures and other correlates of survey outcomes to adjust for non-response: examples from multiple surveys”. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 173.2, pp. 389–407.
- Laaksonen, Seppo (2018). *Survey Methodology and Missing Data*. Cham: Springer.
- Lee, Brian K., Justin Lessler, and Elizabeth A. Stuart (2011). “Weight Trimming and Propensity Score Weighting”. In: *PLoS ONE* 6.3. Ed. by Giuseppe Biondi-Zoccai, e18174. URL: <https://dx.plos.org/10.1371/journal.pone.0018174> (visited on 11/18/2019).
- LeSage, James P. and R. Kelley Pace (2009). *Introduction to Spatial Econometrics*. Boca Raton: Taylor & Francis.
- (2014). “The Biggest Myth in Spatial Econometrics”. In: *Econometrics* 2.4, pp. 217–249.

- Macera, Caroline A. et al. (1990). “Patterns of non-response to a mail survey”. In: *Journal of Clinical Epidemiology* 43.12, pp. 1427–1430.
- Manski, Charles F. (1993). “Identification of Endogenous Social Effects: The Reflection Problem”. In: *Review of Economic Studies* 60.3, pp. 531–542.
- Musterd, Sako (2005). “Social and Ethnic Segregation in Europe: Levels, Causes, and Effects”. In: *Journal of Urban Affairs* 27.3, pp. 331–348.
- Musterd, Sako et al. (2017). “Socioeconomic segregation in European capital cities. Increasing separation between poor and rich”. In: *Urban Geography* 38.7, pp. 1062–1083.
- Olson, Kristen (2013). “Paradata for Nonresponse Adjustment”. In: *The ANNALS of the American Academy of Political and Social Science* 645.1. Ed. by Douglas S. Massey and Roger Tourangeau, pp. 142–170.
- Pace, R. Kelley and James P. LeSage (2010). “Omitted Variable Biases of OLS and Spatial Lag Models”. In: *Progress in Spatial Analysis*. Ed. by Antonio Páez et al. Berlin: Springer Berlin Heidelberg, pp. 17–28.
- Pinheiro, José and Douglas Bates (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.
- Rosenbaum, Paul R. and Donald B. Rubin (1983). “The Central Role of the Propensity Score in Observational Studies for Causal Effects”. In: *Biometrika* 70.1, pp. 41–50.
- Särndal, Carl-Erik and Sixten Lundström (2005). *Estimation in surveys with nonresponse*. Wiley Series in Survey Methodology. Chichester: John Wiley & Sons.
- (2008). “Assessing Auxiliary Vectors for Control of Nonresponse Bias in the Calibration Estimator”. In: *Journal of Official Statistics* 24.2, pp. 167–191.
- Särndal, Erik (2007). “The calibration approach in survey theory and practice”. In: *Survey Methodology* 33.2, pp. 99–119.
- Sax, Linda J., Shannon K. Gilmartin, and Alyssa N. Bryant (2003). “Assessing response rates and nonresponse bias in web and paper surveys”. In: *Research in Higher Education* 44.4, pp. 409–432.
- Schouten, Barry (2007). “A Selection Strategy for Weighting Variables Under a Not-Missing-at-Random Assumption”. In: *Journal of Official Statistics* 23.1, pp. 51–68.

- Skinner, C. J. and D'arrigo (2011). “Inverse probability weighting for clustered non-response”. In: *Biometrika* 98.4, pp. 953–966.
- Valliant, Richard, Jill A Dever, and Frauke Kreuter (2018). *Practical tools for designing and weighting survey samples*. 2nd ed. Statistics for Social and Behavioral Sciences. OCLC: 1035510750. Cham: Springer.
- Vega, Solmaria Halleck and J. Paul Elhorst (2015). “The SLX model”. In: *Journal of Regional Science* 55.3, pp. 339–363.
- Watanabe, Megumi, Kristen Olson, and Christina Falci (2017). “Social isolation, survey nonresponse, and nonresponse bias: An empirical evaluation using social network data within an organization”. In: *Social Science Research* 63, pp. 324–338.
- Zangger, Christoph (2019). “Making a place for space: Using spatial econometrics to model neighborhood effects”. In: *Journal of Urban Affairs* 41.8, pp. 1055–1080.

A. Appendix

Figure A.1: Mean squared error mean outcome, propensity weights ($\text{cor}(oX_1, uX_1) = -0.2$)

Figure A.2: Mean squared error mean outcome, trimmed weights ($\text{cor}(oX_1, uX_1) = -0.2$)