Advance Access publication January 27, 2010

Political Analysis (2010) 18:172–188
doi:10.1093/pan/mpp045

A Selection Model for Panel Data: The Prospects of Green Party Support

Martin Spiess
Department of Psychology, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany
e-mail: martin.spiess@uni-hamburg.de (corresponding author)

Martin Kroh
Socio-Economic Panel Study, German Institute for Economic Research (DIW Berlin), Mohrenstrasse 58, 10117 Berlin, Germany
e-mail: mkroh@diw.de

Although sample selection bias is a frequent problem of applied research, there has been no generalization of sample selection models with binary dependent variables of interest to data with temporal error correlations. We suggest a generalized estimating equation approach to panel data selection models, considering binary responses in both equations. We demonstrate the utility of this model by a simulation study and by analyzing highly unbalanced annual panel data taken from the German Socio-Economic Panel Study covering two decades of Green party support.

1 Introduction

Sample selection is a common problem of applied social science research. Although many studies claim to apply to entire populations, their samples are often subject to coverage problems. In political research, sample selection has been found to operate, for instance, in studies of public opinion (e.g., Bartels 1994; Berinsky 1999), electoral behavior (e.g., Timpone 1998, 2002; Goodliffe 2001), interest groups (e.g., Grier, Munger, and Roberts 1994; McCarty and Rothenberg 1996), party emergence and party behavior (e.g., Jackman and Vella 1991; Hug 2000), comparative research (for an overview see Geddes 1990; Hug 2003), and international relations (e.g., special issue of *International Interactions*, 2002).

Sample selection arises in situations in which the observations are not selected independently from the variables of interest. General population surveys that draw on register data for sampling purposes, for instance, exclude certain strata of the population from sampling (for instance, homeless people) and assign others a lower selection probability (for instance, regionally mobile individuals, who are more difficult to identify in register data). Selectivity may result not only from decisions on sampling design but also from self-selection, that is, the units of analysis may differ in their probability of entering the sample of observations. For instance, respondents at the lower end of the income distribution are less likely to participate in surveys, but when they do participate, they are more likely to refuse to answer on certain income

© The Author 2010. Published by Oxford University Press on behalf of the Society for Political Methodology.
All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

components such as welfare (Fitzgerald, Gottschalk, and Moffitt 1998; Moore, Stinson, and Welniak 1999).

The so-called Heckman model was in fact designed to deal with problems of self-selection and has become one popular solution to the problem. In its original application to labor economics, respondents participating in the labor force were found not to represent a random sample of the overall population; any differences in labor income between groups in the sample may, thus, not necessarily reflect expected wage differences between particular groups in the general population (Heckman 1974). Either estimated by maximum likelihood (ML) or in a two-equation framework, the Heckman model estimates a selection equation in order to obtain information on the concept of interest, whereas the (linear) equation of interest is given for only those units of analysis that were selected or selected themselves into observation. Allowing for a correlation between both processes, the model controls for sample selection bias in the substantive equation due to unobserved variables (Heckman 1976, 1979).

The original Heckman model has been criticized repeatedly for its assumption that the error term in the equation of interest and the error term in the selection equation follow a bivariate normal distribution (cf. Winship and Mare 1992). Some opponents argue that parameter estimates of the Heckman model are inconsistent if the distribution of the error terms is misspecified (Manski, 1989; Newey, 1999; Little and Rubin 2002). The exclusion restriction of the original Heckman model—that is, the necessity of an extra explanatory factor of the selection equation that is strictly exogenous to the equation of interest—has also been subjected to some critique (e.g., Sartori 2003). The original Heckman model has received considerable scholarly attention in the past 30 years countering some of this criticism and has been constantly amended to compensate for selectivity in alternative settings (for an overview, see Winship and Mare 1992; Vella 1998).

Semi- and nonparametric estimation approaches have been developed (e.g., Andrews 1998) and the model has been extended to cases with, for instance, binary dependent variables in the equation of interest and situations with limited information on the selection process. Moreover, in cases of unit nonresponse with no information on the unobserved units available in the survey, Boehmke (2003) proposes a kind of ''plug-in'' ML approach with a cross-sectional selection model using auxiliary information from outside the sample. Sartori (2003) considers a situation in which selection and the response of interest have the same causes, involve similar decisions, and both responses are close together in time; hence, the sets of covariates are identical and the error terms in the underlying utility equations are very similar, with a correlation equal to or close to one or minus one.

Despite the commonly acknowledged benefits of the selection model and the growing body of literature expanding its original scope, surprisingly, little research has been done on panel data selection models. This may be due to the computational burden imposed by the requirement to evaluate high-dimensional integrals for ML estimation and the common belief that fixed-effects estimation eliminates unobserved heterogeneity and, thus, selectivity (see, e.g., Vella 1998). The latter, however, although true for certain forms of selection bias, does not hold in general (e.g., Kyriazidou 1997). Similarly, Vella (1998) shows that estimating a random effects model does not eliminate selection effects.

To avoid the evaluation of high-dimensional integrals, a general problem in (partly) nonlinear panel and (partly) nonlinear multivariate models (non-ML), two- or three-step differencing approaches with linear equations of interest have been proposed in the context of panel data selection models (for a review, see Vella 1998). Kyriazidou (1997), for instance, proposes a two-step approach to estimate a differenced fixed-effects model with a linear equation of interest in the case of two panel waves. The approach is based on weighted estimation of the equation of interest, where the ''kernel'' weights depend on

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

the estimated effect of selection into the equation of interest. The estimation approach does not rely on strong distributional assumptions, but the estimator seems to be quite sensitive to the choice of a bandwidth parameter. Alternative differencing approaches have been proposed by Rochina-Barrachina (1999) and Gayle and Viauroux (2007), the latter allowing a semiparametric estimation of a dynamic panel selection model in three steps.

A general restriction of these panel selection models is that their model of interest is a linear model for continuous dependent variables. Moreover, the differencing approach in general does not allow the estimation of effects of time-invariant covariates, and its reliance on pairwise differences often leads to the exclusion of cases that are observed only once. Valid inferences are therefore only possible under restrictive assumptions with respect to the excluded cases. In the model by Kyriazidou (1997), for instance, those units with estimated selection effects at two different time points that are not close together receive small weights and are asymptotically ignored, leading to a loss of information depending on the behavior of the units (for a discussion of the differencing approach to panel selection models, see Dustmann and Rochina-Barrachina 2007).

We suggest an alternative estimation strategy to panel data selection models that is highly flexible and covers various panel settings. More specifically, we adopt a generalized estimating equations (GEE) approach (Liang and Zeger 1986) that allows us to consider binary variables in the equation of interest, highly unbalanced data sets with many waves, and the incorporation of arbitrary correlation structures of the selection equation and the equation of interest. Furthermore, our approach allows us to estimate the effects of time-invariant covariates. The price to pay for this flexibility is the need to correctly specify uni- and bivariate distributions. A general introduction into the GEE approach with applications from political research can be found in Zorn (2001). To illustrate our estimation strategy, we draw on empirical data emanating from a long-running panel study and try to answer a question of political interest, namely whether Green parties in Western societies are supported by a single protest generation only or by several recent political generations.

2 The Greens Threatened by Extinction?

In the 1970s, scholars of public opinion identified a general trend in many Western societies toward what was termed ''new politics'' (Inglehart 1971; Baker, Dalton, and Hildebrandt 1981). Since then, however, the research has examined the question of whether this change in public attitudes is actually being sustained (Bürklin 1987; Kitschelt 1988).

One prominent thesis on the subject states that new politics and related phenomena, such as new political movements (e.g., Greenpeace, peace and civil liberty movements, etc.) and new political parties (i.e., mainly environmentalist parties) are a function of the rising standard of living and educational levels. The enduring economic security of those born in the last few decades in Western societies changes their value priorities from material well-being to more idealistic policy goals (Inglehart 1971). Thus, one could conclude that the continuing trend toward increased prosperity would lead to increased support for new politics and new political parties among recent generations as well.

The alternative account holds that new politics were a consequence of the turbulent period of the late 1960s to the late 1970s (Jennings 1987; Hulsberg 1988). Thus, the primary supporters of new politics would be those birth cohorts socialized during that period. Although these cohorts replaced former cohorts when they entered the electorate, they will at some point be replaced by successor cohorts (Bürklin and Dalton 1994). In other words, from this point of view, new politics is a temporal phenomenon carried solely by a single political generation.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

The debate on the ''life expectancy'' of new politics and, thus, Green parties (Kitschelt 1988) has been sparked particularly by longitudinal studies that show a Green electorate that has been aging increasingly since the party's formation in the early 1980s (Bürklin and Dalton 1994). Subsequent research focused on identifying the importance of cohort differences for this trend (Klein and Arzheimer 1997; Kohler 1998). This article seeks to contribute to this debate by identifying the varying support for Green parties while controlling for the declining loyalty of recent cohorts to parties in general.

3 The Panel Data Selection Model

The most common measures of party and candidate support—voting behavior, voting intention, and party identification—are inherently subject to selection bias if one is interested in political support in the entire population (cf. van der Eijk et al. 2006).

Voting behavior, that is, reported party or candidate choice in the previous election, applies to only those respondents who participated in the election. Interviewees who were either not eligible to vote, not registered, or decided to abstain from voting do not reveal their political preferences in surveys of voting behavior. Voting intention, that is, projected party or candidate choice in the upcoming election, is also likely to produce a selective sample with regard to the outcome variable of general party support. Beyond restrictions that also apply to voting behavior, we may encounter respondents who, at the time of the interview, had simply not yet decided which party or candidate to vote for. Finally, party identification is limited to those respondents who conceive of themselves as longstanding party loyals. The party identification survey measure may therefore not reveal the party support of respondents who consider themselves politically unaligned. In panel data, moreover, the fraction of respondents with missing data on the outcome variable on at least one occasion arguably grows with the length of the period under observation.

In this article, we consider repeated observations of individual party identification. The German Socio-Economic Panel Study (SOEP) has one of the longest time series of annual observations on party identification of any panel study (for applications of the data, see Zuckerman 2005; Schmitt-Beck, Weick, and Christoph 2006; Zuckerman and Kroh 2006; Kroh and Selb 2009). Established in 1984 in West Germany with regular refreshment samples in the following years (e.g., adding East Germans in 1990), the ongoing annual survey currently consists of a representative national sample of more than 20,000 individuals in more than 10,000 households (Kroh and Spiess 2008).

The SOEP uses the standard survey instrument for measuring party identification in multiparty contexts, which consists of two questions. First, respondents are asked: ''Many people in Germany lean toward a particular political party, although they occasionally vote for a different party. What about you: Do you lean—generally speaking—toward a particular party?'' Only those who respond ''Yes'' are then asked, ''Which one?''. The response to the first question is the binary selection indicator, y_{it1}, obtained on each of N respondents at each of T points in time ($i = 1, \ldots, N$; $t = 1, \ldots, T$). The response to the second question, party support for the Greens versus any other party, is the binary outcome of primary scientific interest, y_{it2}, which is only observed if $y_{it1} = 1$. In addition, there is a vector of fixed and always observed covariates $\mathbf{x}_{it1}$ thought to be related to y_{it1} and a vector of fixed and always observed covariates $\mathbf{x}_{it2}$ thought to be related to y_{it2}. These covariates are, for instance, indicators of the year of interview and respondents' birth cohort. Both, $\mathbf{x}_{it1}$ and $\mathbf{x}_{it2}$, include a vector of ones.

The model assumes that each outcome is related to a continuous latent variable, y^*_{it1} and y^*_{it2}, respectively. In particular,

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

$$y_{it1} = \begin{cases} 1 & \text{if } c_1 < y^*_{it1} \\ 0 & \text{else} \end{cases} \quad \text{and} \quad y_{it2} = \begin{cases} 1 & \text{if } c_2 < y^*_{it2} \\ 0 & \text{else} \end{cases}.$$

The latent model is

$$y^*_{itj} = \eta_{itj} + \epsilon_{itj} \quad \text{and} \quad \eta_{itj} = \mathbf{x}^T_{itj}\boldsymbol{\beta}_j,$$

where $j = 1$ denotes the selection equation and $j = 2$ denotes the equation of primary interest, $\boldsymbol{\beta}_j$ is an unknown equation-specific vector of parameters of the mean structure. Including a constant term in the model equations, the unknown thresholds c_1 and c_2 are not identified. Thus, following usual convention, we impose the restrictions $c_1 = c_2 = 0$.

If, conditional on the covariates, Green party support, y^*_{it2}, were unrelated to general partisanship, y^*_{it1}, a strategy that estimates a party support regression on the basis of the observed cases would lead to valid inferences. A large body of research suggests, however, that—with a given set of covariates—voting behavior, voting intention, and party identification $\left(y^*_{it2}\right)$ are associated with reported voter turnout, intended turnout, and general partisanship $\left(y^*_{it1}\right)$, respectively (Dubin and Rivers 1990).[1]

Teixeira (1992) reviews the literature on turnout and concludes that nonvoters hold more liberal political preferences than the active electorate; Abramson, Aldrich, and Rohde (1995) show that some parties are better able to systematically mobilize their party constituency at an early stage of the campaign than others, and Richardson (1991) demonstrates that traditional cleavage parties attract particularly stable party support. Parties that reflect established sociostructural conflicts to a lesser extent, such as the German Greens, experience less durable support from their constituency. Also with respect to core explanatory variables of our application—the support for the Green party in a longitudinal perspective—a sample of highly mobilized respondents may lead to biased conclusions as mobilization has been shown to vary systematically by age (e.g., Verba and Nie 1972), cohort (e.g., Dalton and Wattenberg 2000; Franklin 2004), and period (e.g., Beck and Jennings 1979; Rosenstone and Hansen 1993).

Moreover, the two-fold character of new politics may dispose the self-selection of Green party supporters with respect to the measure of party identification. New politics not only describes a certain policy agenda, including peace, environmentalism, and equality but also a certain style of political involvement. Most Green parties emerged from thriving grassroots democratic movements and defined themselves as nontraditional organizations in which electoral success is of secondary importance. The German Green party represents a typical example of such a new political movement (Müller-Rommel 1985). Indeed, Petra Kelly, one of the leaders of the Green party in Germany, coined the notion of the Greens as an ''anti-party party.'' One would therefore expect loyalty toward a parliamentary party organization to stand at odds with (latent) Green party support. Or, in other words, we expect Green party support, y^*_{it2}, to be negatively associated with general partisanship, y^*_{it1}, conditional on observed covariates.

Let $\boldsymbol{\epsilon}_{it}$ be the (2×1) vector with elements ϵ_{it1}, ϵ_{it2}, and $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}'_{i1}, \ldots, \boldsymbol{\epsilon}'_{iT})'$. The presumed negative association of y^*_{it2} and y^*_{it1}, given the covariates $\mathbf{x}_{it1}$ and $\mathbf{x}_{it2}$, amounts to a negative association of ϵ_{it1} and ϵ_{it2}. Throughout we will assume that $[\boldsymbol{\epsilon}_i|(\mathbf{x}_{i11}, \ldots, \mathbf{x}_{iT2})]$ is

[1]The endogeneity of turnout in models of party and candidate preference is most explicitly acknowledged by rational choice theories. Already Downs (1957) defines turnout as a function of the party differential, that is the expected differences in party utilities.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

identically and independently distributed (iid) with $E(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}$ for all i. Since the GEE approach discussed below does not involve higher order specifications, for our panel selection model, only the bivariate distributions of $\boldsymbol{\epsilon}_{it}$ need to be correctly specified, and these will be assumed to be bivariate normal. This is in contrast to ML estimation, which would require the specification of a high-dimensional distribution. Thus, at the price of a loss of efficiency, the estimating approach discussed below is less vulnerable to violations of assumptions about higher order moments. However, in general, the gain in robustness is far greater than the loss in efficiency (Liang and Zeger 1995).

4 The GEE Approach

4.1 *The Estimating Equations*

For simplicity, we will first take a cross-sectional view and then generalize ideas to the panel case. With the above assumptions, the latent responses y^*_{it1} are normally distributed and the (marginal) probability $\Pr(y_{it1} = 1|\mathbf{x}_{it1})$, which is equal to $E(y_{it1}|\mathbf{x}_{it1})$, is given by the value of the cumulative standard normal distribution function $\Phi(\cdot)$ at $(\mathbf{x}_{it1}\boldsymbol{\beta}_1)/\sigma_1$, denoted as μ_{it1}. Thus, our model for y_{it1} ($i = 1, \ldots, N$, $t = 1, \ldots, T$) is a simple probit model. Since not all the remaining parameters are identified, we impose the additional restriction $\sigma_1^2 = \mathrm{var}(\boldsymbol{\epsilon}_{it1}) = 1 \quad \text{for all} \quad i, t.$

However, the conditional distribution of $\left[y^*_{it2}\middle|\mathbf{x}_{it2}, y_{it1} = 1\right]$ is only normal if $f\left(y^*_{it2}\middle|\mathbf{x}_{it2}, y_{it1}\right) = f\left(y^*_{it2}\middle|\mathbf{x}_{it1}\right)$, where $f(\cdot)$ is the (conditional) density function of $\left[y^*_{it2}\middle|\mathbf{x}_{it2}, y_{it1}\right]$. Hence, one has to correct for the possible selection bias induced by the dependence of $y^*_{it2}\middle|\mathbf{x}_{it2}$ on y_{it1}. The ML method is one alternative to Heckman's well-known two-step estimation approach (e.g., Heckman 1979) that has been proposed in the case of cross-sectional data and a linear substantive model equation. This is the approach we adopt for the cross-sectional case. Looking more closely at the score equations in this case reveals that, besides some weak regularity conditions, we only need to correctly specify the (conditional) means of the dependent variables y_{it1} and y_{it2}.

Since instead of y^*_{it2} only the binary indicator y_{it2} is observed, we need to find the conditional probability $\Pr(y_{it2} = 1|\mathbf{x}_{it2}, y_{it1} = 1)$, which is

$$\Pr(y_{it2} = 1,\ \ y_{it1} = 1|\mathbf{x}_{it1},\ \ \mathbf{x}_{it2})/\Pr(y_{it1} = \mathbf{x}_{it1}),$$

where we presuppose that $\Pr(y_{it1} = 1|\mathbf{x}_{it1}, \mathbf{x}_{it2}) = \Pr(y_{it1} = 1|\mathbf{x}_{it1})$ and $\Pr(y_{it2} = 1|\mathbf{x}_{it1}, \mathbf{x}_{it2}, y_{it1} = 1) = \Pr(y_{it2} = 1|\mathbf{x}_{it2}, y_{it1} = 1)$. Assuming bivariate normality, $\Pr(y_{it2} = 1, y_{it1} = 1|\mathbf{x}_{it1}, \mathbf{x}_{it2})$ is the value of the cumulative bivariate standard normal distribution function at $(\mathbf{x}'_{it1}\boldsymbol{\beta}_1)/\sigma_1$, $(\mathbf{x}'_{it2}\boldsymbol{\beta}_2)/\sigma_2$ and ρ_t, where ρ_t is the correlation between $\boldsymbol{\epsilon}_{it1}$ and $\boldsymbol{\epsilon}_{it2}$ which accounts for the selection on unobserved variables. The probability $\Pr(y_{it2} = 1|\mathbf{x}_{it2}, y_{it1} = 1)$, which is equal to $E(y_{it2}|\mathbf{x}_{it2}, y_{it1} = 1)$, will be denoted as $\mu_{it2|it1}$. Since again not all parameters are identified, we impose the additional restriction

$$\sigma_2^2 = \mathrm{var}(\boldsymbol{\epsilon}_{it2}) = 1 \quad \text{for all} \quad i, t.$$

If $T = 1$, the log-likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \rho_t)'$ based on the observed data is thus

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{N} \ln \mu_{it1}^{y_{it1}}(1-\mu_{it1})^{1-y_{it1}}\left[\mu_{it2|it1}^{y_{it2}}(1-\mu_{it2|it1})^{1-y_{it2}}\right]^{y_{it1}}. \tag{1}$$

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

The score equations for $\boldsymbol{\theta}$ can be written as

$$\mathbf{0} = \sum_{i=1}^{N} \mathbf{X}_{\mathrm{it}}^{'} \mathbf{D}_{\mathrm{it}}^{'} \mathbf{W}_{it} \mathbf{e}_{it}, \tag{2}$$

where

$$\begin{aligned}
&\mathbf{X}_{it} = \mathrm{diag}(\mathbf{x}_{\mathrm{it1}}^{'}, \mathbf{x}_{\mathrm{it2}}^{'}, 1)\\
&\mathbf{e}_{it} = (\mathbf{y}_{it} - \boldsymbol{\mu}_{it}), \quad \text{where } \mathbf{y}_{it} = (y_{it1}, y_{it1}\ y_{it2})^{'}, \ \boldsymbol{\mu}_{it} = (\mu_{it1}, y_{it1}\ \mu_{it2|it1})^{'}\\
&\mathbf{D}_{it} = \tfrac{\partial \boldsymbol{\mu}_{it}}{\partial \boldsymbol{\eta}_{it}^{*}}, \quad \text{where} \quad \boldsymbol{\eta}_{it}^{*} = (\eta_{it1}, \eta_{it2}, \rho_t)^{'}, \quad \text{and}\\
&\mathbf{W}_{it} = \mathrm{diag}([\mu_{it1}(1-\mu_{it1})]^{-1}, \ [\mu_{it2|it1}(1-\mu_{it2|it1})]^{-1}).
\end{aligned}$$

Premultiplying the second diagonal element in $\mathbf{W}_{it}$ with y_{it1} accounts for the fact that y_{it2} is only observed if $y_{it1} = 1$. Note that equation (2) can be interpreted as GEEs introduced by Liang and Zeger (1986), where a crucial condition for the asymptotic properties to hold is that, conditional on the covariates, the mean model, $\boldsymbol{\mu}_{it}$, is correctly specified. In fact, taking the mean of $(y_{it1} - \mu_{it1})$ and $(y_{it1}y_{it2} - y_{it1}\mu_{it2|it1})$, we ensure that $E(\mathbf{e}_{it}) = \mathbf{0}$ if $\Pr(y_{it1} = 1|\mathbf{x}_{it1})$ and $\Pr(y_{it2} = 1, y_{it1} = 1|\mathbf{x}_{it1}, \mathbf{x}_{it2})$ are correctly specified. Pannenberg and Spiess (2009) show that the same estimating equations could have been derived if the generalized method of moments approach had been adopted (for the generalized methods of moments [GMM] approach, see Hansen 1982; Avery, Hansen, and Hotz 1983).

With a little additional notation, the extension to the panel case with $T > 1$ is straightforward. Let $\mathbf{1}_T$ be a $(T \times 1)$-dimensional vector of ones and $\mathbf{e}_i$ be the $(2T \times 1)$ vector with T-dimensional stacked vectors $\mathbf{e}_{i1} = (\mathbf{y}_{i1} - \boldsymbol{\mu}_{i1})$ and $\mathbf{e}_{i2|i1} = (\mathbf{y}_{i2} - \boldsymbol{\mu}_{i2|i1})$ comprising elements y_{it1}, μ_{it1}, y_{it2}, and $\mu_{it2|it1}$, $t = 1, \ldots, T$, respectively. Since it only depends on y_{it1} whether or not y_{it2} is observed, the marginal distributions of y_{it1} and y_{it2} remain the same as in the cross-sectional case. Thus, adopting the GEE approach, the estimating equations for $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^{'}, \boldsymbol{\beta}_2^{'}, \rho_t)'$, restricting the correlations ρ_t to be equal over time, can be written as

$$\mathbf{0} = \sum_{i=1}^{N} \mathbf{X}_i^{'} \mathbf{D}_i^{'} \mathbf{W}_i \mathbf{e}_i, \tag{3}$$

where

$$\begin{aligned}
&\mathbf{X}_i = \mathrm{diag}(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \mathbf{1}_T),\\
&\mathbf{D}_i = \begin{pmatrix} \mathbf{D}_{11i} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_{21i} & \mathbf{D}_{22i} & \mathbf{D}_{23i} \end{pmatrix},\\
&\mathbf{D}_{11i} = \tfrac{\partial \boldsymbol{\mu}_{i1}}{\partial \boldsymbol{\eta}_{i1}} \quad \mathbf{D}_{21i} = \tfrac{\partial \boldsymbol{\mu}_{i2|i1}}{\partial \boldsymbol{\eta}_{i1}} \quad \mathbf{D}_{22i} = \tfrac{\partial \boldsymbol{\mu}_{i2|i1}}{\partial \boldsymbol{\eta}_{i2}}\\
&\mathbf{D}_{23i} = \tfrac{\partial \boldsymbol{\mu}_{i2|i1}}{\partial \boldsymbol{\rho}} \quad \boldsymbol{\rho} = (\rho_1, \ldots, \rho_T)^{'},\\
&\boldsymbol{\eta}_{i1} = (\eta_{i11}, \ldots, \eta_{iT1})^{'} \quad \text{and} \quad \boldsymbol{\eta}_{i2} = (\eta_{i12}, \ldots, \eta_{iT2})^{'},\\
&\mathbf{W}_i = \mathrm{diag}\big(\mathbf{1}_T, \mathbf{y}_{i1}\big)\mathbf{V}_i^{-1/2}\mathbf{R}^{-1}\mathbf{V}_i^{-1/2}\ \mathrm{diag}\big(\mathbf{1}_T, \mathbf{y}_{i1}\big),\\
&\mathbf{V}_i = \mathrm{diag}(\boldsymbol{\mu}_{i1}(1-\boldsymbol{\mu}_{i1}), \ \boldsymbol{\mu}_{i2|i1}(1-\boldsymbol{\mu}_{i2|i1})).
\end{aligned}$$

and $\mathbf{R}$ is a $(2T \times 2T)$-"working" correlation matrix to model dependencies in the residuals $\mathbf{e}_i$ over time. Although the structure of $\mathbf{R}$ is usually unknown, equation (2) implies that the diagonal elements of the $(T \times T)$ off-diagonal blocks are equal to zero. Note, however, that

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

$\boldsymbol{\theta}$ can consistently be estimated even if the $(T \times T)$ diagonal blocks in $\mathbf{R}$ are misspecified, although efficiency is lost in this case (e.g., Liang and Zeger 1986). Further note that the estimating equations (3) are not equal to the score equations derived from a log-likelihood function.

4.2 *Estimation of the Model*

Estimating equations (3) must be solved numerically. Thus, the vector of estimates, $\hat{\boldsymbol{\theta}}$, is iteratively calculated with updated value in the $(k + 1)$th iteration given by

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{D}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{X}_i \right)^{-1}_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_k} \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{D}_i' \mathbf{W}_i \mathbf{e}_i \right)_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_k}.$$

Note that $\mathbf{W}_i$ involves the inversion of the $(2T \times 2T)$ working correlation matrix $\mathbf{R}$, in our application a maximum of 14 interviews on party identification between 1985 and 2007 a (28×28) matrix. However, from equation (2), it is clear that the diagonal elements of the $(T \times T)$ off-diagonal block in $\mathbf{R}$ must be zeros. Since additionally the diagonal blocks in $\mathbf{R}$ are the working correlation matrices of the selection indicators and the selection indicators times, the outcomes of interest, respectively, it seems hard to specify a common structure for $\mathbf{R}$. Thus, we restrict $\mathbf{R}$ to be a block diagonal matrix, consisting of two $(T \times T)$ working correlation matrices $\mathbf{R}_1$ and $\mathbf{R}_2$. In principle, two different structures could be assumed for $\mathbf{R}_1$ and $\mathbf{R}_2$. We will assume the same structure for both, however, allowing for different values of the correlation structure parameters. The assumed structure is of a Toeplitz form, where the correlations depend only on the time distance without further restrictions, amounting to the estimation of 2×13 correlation structure parameters. The Toeplitz structure is still flexible, but the chance of ending up with a nonpositive definite working correlation matrix is much lower than with an unstructured correlation matrix.

Solving equation (3) leads to estimators $\hat{\boldsymbol{\theta}}$ that are consistent for the ''true'' parameter $\boldsymbol{\theta}_0$ and are asymptotically normal under mild regularity conditions (Liang and Zeger 1986). Its asymptotic covariance matrix can consistently be estimated by

$$\widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}}) = \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{D}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{X}_i \right)^{-1} \times \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{D}_i' \mathbf{W}_i \mathbf{e}_i \mathbf{e}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{X}_i \right) \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{D}_i' \mathbf{W}_i \mathbf{D}_i \mathbf{X}_i \right)^{-1},$$

where all unknowns are replaced by their estimates. A simulation study, the results of which are reported in Appendix, implies that the proposed approach performs well in different settings.

5 Results

Table 1 reports the percent of valid votes received by the Green party in elections to the national assembly since 1983, when they passed the 5% legal threshold for the first time. The election results of the Greens range between 4.8% and 9.4% of the valid votes. During that period, turnout at elections to the national assembly hovered at around 80%. Note that the data on the post-unification period pertain to West Germany only (including West Berlin).

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Table 1 Green Party support in West Germany, 1983–2007

	Election results (Bundestag)		*Survey data (SOEP)*	
	Turnout rate	*Green votes*	*Partisan rate*	*Green support*
1983	.884	.056	—	—
1984	—	—	.629	.060
1985	—	—	.611	.064
1986	—	—	.610	.065
1987	.831	.083	.662	.072
1988	—	—	.630	.061
1989	—	—	.615	.067
1990	—	—	.629	.052
1991	.786	.048	.633	.049
1992	—	—	.544	.057
1993	—	—	.505	.081
1994	.794	.079	.534	.074
1995	—	—	.568	.104
1996	—	—	.545	.108
1997	—	—	.534	.100
1998	.827	.073	.549	.093
1999	—	—	.566	.066
2000	—	—	.503	.077
2001	—	—	.485	.081
2002	.806	.094	.497	.076
2003	—	—	.530	.103
2004	—	—	.451	.111
2005	.784	.088	.477	.122
2006	—	—	.514	.106
2007	—	—	.439	.110

Note. All entries pertain to the West German electorate including West Berlin.

The proportion of respondents in the SOEP who report loyalty to any party has declined gradually over time from more than 60% in the 1980s to below 50% today, a finding that is in line with previous research (Dalton and Wattenberg 2000). At the same time, among those who declared themselves to have party loyalties, support for the Greens relative to all other parties has ranged between 4.9% and 12.2% across the same period.

Two strategies suggest themselves for analyzing Green party support on the basis of the party identification measure. Either one defines the incidence of a respondent reporting party loyalty in general and party loyalty to the Greens in particular as an indicator of Green party support. All other response patterns denote no support for the Greens, including interviewees who report no persistent party leanings and interviewees who report persistent leaning toward another party than the Greens. This approach has been adopted by Kohler (1998), for instance. The alternative approach to analyzing Green party support in the entire sample would be to consider only the specific party preference y_{it2} as the relevant indicator and the general partisanship question y_{it1} as only the selection process that elicits an individual's report of party support or lack thereof. In other words, the latter approach assumes that respondents who declare not being longstanding party loyals are nonetheless able to make a choice between parties. The following sections will contrast the empirical results of both strategies of analysis. Note, however, that given their respective definitions

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Table 2 GEE regression models of Green Party support, 1985–2007

	Probit panel		*Probit panel selection*			
	Model 1		*Model 2*			
	β	*SE*	γ	*SE*	β	*SE*
Age groups (years)						
<30	—	—	—	—	—	—
30–45	−.064	(.017)***	.120	(.020)***	−.066	(.036)*
45–60	−.107	(.026)***	.104	(.031)***	−.079	(.054)
60+	−.107	(.040)***	.131	(.042)***	−.080	(.082)
Political generations						
Adenauer (CDU)	—	—	—	—	—	—
Erhard (CDU)	.275	(.030)***	−.009	(.036)	.307	(.103)***
Brandt (SPD)	.682	(.035)***	−.074	(.034)**	.767	(.090)***
Schmidt (SPD)	.905	(.037)***	−.178	(.037)***	1.151	(.093)***
Kohl (CDU)	.751	(.040)***	−.299	(.041)***	1.110	(.099)***
Schröder (SPD)	.479	(.047)***	−.357	(.053)***	1.059	(.118)***
Year of interview (all 15 parameters omitted)						
Interest in politics	—	—	.479	(.008)***	—	—
Political engagement	—	—	.109	(.016)***	—	—
Civic engagement	—	—	.081	(.012)***	—	—
Organization membership	—	—	.129	(.019)***	—	—
Distance to election	—	—	−.023	(.012)*	—	—
Constant	−2.167	(.044)***	−.611	(.091)***	−1.969	(.105)***
ρ			−.178	(.050)***		

Source: SOEP.
Note. Number of observations = 76,768, N = 14,444. CDU: Christian Democrat; SPD: Social-Democrat.
*p < .10; **p < .05; ***p < .01.

of Green party support, both approaches produce ''valid'' results. Although the first approach is more informative about the development of Green strongholds in the electorate, the second approach is arguably more informative about the electoral prospects of the Greens in general as it explicitly disentangles processes of political detachment from trends in the electoral support of specific parties.

The phenomenon of new politics and support for Green parties is said to primarily apply to established Western democracies (Inglehart 1971). This article therefore considers only those birth cohorts socialized in the former West Germany and, thus, excludes both those respondents who participated in German elections prior to the first free election in West Germany in 1948 and those who participated in elections to the East German parliament. Moreover, the analysis omits all respondents with a migration background who participated in elections to offices in their country of origin. In all these cases, cohort membership may not act as an omnibus proxy for shared political experiences in the West German political system (Mannheim 1928). Thus, we end up with a sample of N = 14,440 individuals responding at least once in the (T = 14) panel waves between 1985 and 2007 considered here, leading to 76,768 observations.

Model 1 of Table 2 analyzes longstanding Green party support. That is, nonpartisans ($y_{it1} = 0$) and partisans of other parties ($y_{it1} = 1$, $y_{it2} = 0$) are coded zero, and those who report persistent leaning toward the Greens ($y_{it1} = 1$, $y_{it2} = 1$) are coded one. Disentangling cohort, age, and period effects, the unbalanced panel model estimates annual parameters of

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

the date of interview (omitted in Table 2), the contrast effects of four age groups, and the contrast effects of six generational groups. The generational variable classifies respondents according to their first participation in a national election: if respondents entered the electorate, for instance, during the chancellorship of Gerhard Schröder of the Social Democratic Party in the 1998 or 2002 Bundestag elections, they are categorized as the ''Schröder Generation.''

The estimates of Model 1 suggest that once period effects are controlled for, the aging of respondents is associated with declining support for the Greens. Respondents below 30 years have the highest likelihood of reporting party identification with the Green party. However, generational differences are more pronounced than age differences. The ''Schmidt Generation'' entering the electorate in the second half of the 1970s seems to be a stronghold of Green party support. Earlier but also later political generations show much lower levels of durable Green party support. Estimates of Model 1 may thus be interpreted as indicative of the validity of the hypothesis that new politics and accompanying political phenomena like Green parties are temporary in nature, carried only by certain political generations. Moreover, within these generations, support for the Greens declines by age. Model 1, which compares longstanding supporters of the Greens with all other respondents, suggests that a gradual ''graying'' of the Greens will take place in the foreseeable future.

Model 2 of Table 2 reports regression estimates of support for the German Greens while controlling for possible selectivity of general partisanship. A first selection equation regresses the probability of reporting general political inclination on age groups, political generations, period, and on additional covariates. These additional explanatory variables are included in the model to avoid identification problems and to improve the prediction of the selection process and, thus, to increase the precision of estimates in the substantive equation. The choice of these covariates was made on basis of the literature on the formation of political attachments (e.g., Campbell et al. 1960; Fiorina 2002; Green, Palmquist, and Schickler 2002; Zuckerman 2005). Note that for identification reasons, the selection equation needs to include extra covariates that are exogenous of the substantive equation (Sartori 2003). That is, among the large number of possible predictors of the respondents' willingness to declare a party identification, we need to specify at least one explanatory variable that is independent of the choice for the Green versus another party. To meet this exclusion restriction, we specified rather general indicators of political involvement in the selection equation that can plausibly be expected to be independent of political color: interest in politics, involvement in local politics, civic engagement, membership in a social organization, and temporal distance from the national parliamentary elections.[2] Each of these indicators of political involvement affects respondents' willingness to declare a party identification in the predicted direction: partisans are politically more interested and active; they are more often members of organizations and are more active in civic groups. Moreover, the closer the date of an election, the more likely that respondents will report partisanship. The γ-estimates pertaining to the selection equation also suggest increased rates of partisanship in respondents above 30 and particularly above 60 years of age and an increasing detachment of each consecutive generation from parties in general. A person socialized during the chancellorships of Adenauer or Erhard—a period dominated by the Christian Democrats and characterized by post-World War II

[2]That is, the minimal log transformed absolute distance of the date of the interview from elections to the Bundestag.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Table 3 Working Toeplitz correlation matrices for the selection equations (upper triangular) and the equations of interest (lower triangular), for $t = 1, \ldots, 9$

		T								
		1	2	3	4	5	6	7	8	9
t	1	—	.549	.500	.498	.456	.446	.425	.412	.388
	2	.775	—	.549	.500	.498	.456	.446	.425	.412
	3	.746	.775	—	.549	.500	.498	.456	.446	.425
	4	.714	.746	.775	—	.549	.500	.498	.456	.446
	5	.662	.714	.746	.775	—	.549	.500	.498	.456
	6	.645	.662	.714	.746	.775	—	.549	.500	.498
	7	.617	.645	.662	.714	.746	.775	—	.549	.500
	8	.603	.617	.645	.662	.714	.746	.775	—	.549
	9	.558	.603	.617	.645	.662	.714	.746	.775	—

reconstruction—has, ceteris paribus, the highest likelihood of entering the substantive equation and, thus, the highest chances of reporting party preferences.

Restricting $\rho_t = \rho$ for all t = 1,...,T, the estimated correlation between residuals of the selection equation and substantive equation, $\hat{\rho}$, is −.178. With a SE of .050, the estimate is well beyond accepted levels of statistical significance and suggests a negative association between the latent propensity of individuals to report general partisanship and a propensity to support the Green party relative to any other party. After controlling for selectivity, β-estimates from Model 2 suggest a different story than estimates of Model 1. Although we again find generational differences in Green party support, these are somewhat less pronounced than in Model 1, and more importantly, for Green party strategists geared toward long-term electoral success, the support level for the Greens relative to other parties is relatively stable among the most recent Schmidt, Kohl, and Schröder Generations. Estimates of Model 2 may thus be interpreted as indicative of the validity of the hypothesis that new politics and accompanying phenomena such as Green parties are durable in nature and carried by political generations raised during times of continuous economic stability and a high standard of living. Yet, the same factors seem also to be associated with a declining propensity of voters to consider themselves as longstanding party loyals.

Finally, Table 3 gives exemplary results on the working correlation of the selection model for nine time points. Recall that we used a Toeplitz structure for the correlation over time in both selection and substantive equations. The working correlation in the selection equation over time is .55 for the first two annual lags and for the substantive equation .78. It drops to .39 in the selection equation and .56 in the substantive equation for the time lags of order 9. Note that these correlations are based on the observed residuals and not on the latent error term and are therefore known to be attenuated with respect to the latter. That is, the stability of respondents' inclination to consider themselves party loyals and their inclination to support the Greens relative to any other party is likely to be even higher than indicated by the correlations over time.

6 Conclusions

Sample selection is a common problem of applied research. By way of sampling deficiencies and self-selection, relevant units often do not enter the equation of interest.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Beyond the traditional applications of selection models in labor market studies (Heckman 1976, 1979), such models have proven valuable tools of analysis in practically all fields of political research. Although the benefits of sample selection modeling are widely recognized, to the best of our knowledge, no straightforward generalization of this approach to the case of panel data with a binary dependent variable in the equation of interest has been proposed. We suggest such an extension using a GEE framework.

Our approach does not require the evaluation of high-dimensional integrals but nevertheless allows the incorporation of any (positive definite) correlation structure over time, that is, it is not restricted to random effects models with one or two effects. Further, it is easily implemented even for unbalanced models or many observation points in time, and weighted versions to account for nonresponse or informative sampling are easily available. Furthermore, the estimating equations can easily be adapted to allow ordered or unordered categorical variables or count data as dependent variables in the equation of interest. It also can be fruitfully generalized in two ways: First, misspecifications due to distributional assumptions can be avoided by semi-nonparametric generalizations to model the mean. Second, the approach can be generalized to incorporate lagged dependent variables following the literature on GMM estimation.

It should be noted that the proposed approach is not based on unobserved fixed effects. Thus, the consistency of the estimators depends on the assumption that the errors are uncorrelated with the covariates. The inclusion of instrumental variables has been shown to solve this problem in the GMM context (e.g., Wooldridge 2002) and is in principle applicable in the GEE context as well. A topic that has received a considerable amount of attention is the fact that falsely assuming bivariate normality in cross-sectional panel models may lead to severely biased estimators. However, in the case of linear models for the variable of scientific interest, simulation studies suggest that the bias may not be large (e.g., van der Klaauw and Koning, 2003), which is not surprising, since the estimators of the regression parameters are consistent regardless of the distributional assumption. This can, however, be expected to be different in the binary model. Thus, in further work, the semi-nonparametric of van der Klaauw and Koning (2003) could in principle be adopted in the context of the proposed panel selection model. However, estimation of these models is cumbersome and may even be unstable since essential aspects of the unknown distribution have to be estimated in addition to the unknown parameters.

Appendix: A Simulation Study

To illustrate the merits of our panel selection model, we perform a small simulation study. The data sets are simulated according to the model described in Section 3 with $T = 5$ and four covariates generated independently of each other. The covariates follow a standard normal, a Bernoulli, a uniform, and a Poisson distribution. All covariates had an effect in the selection equation with true regression parameter values $\gamma_1 = .6$, $\gamma_2 = -.7$, $\gamma_3 = -.4$, and $\gamma_4 = .01$. Only the standard normally distributed and the Bernoulli variable had an effect in the equation of substantive interest, with true regression parameter values $\beta_1 = .5$ and $\beta_2 = .3$. In the data-generating process, we put the probability of self-selection into the equation of interest at .4.

Moreover, to demonstrate the robustness of the panel selection model, we allow for attrition. We generate unit nonresponse based on a missing-at-random mechanism, where the probability of response depends only on the covariates included in the selection model

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

and on one additional uniformly distributed variable, which was generated independently from all other variables. In addition, we simulated the new entrance of units into the sample with a probability of .2. Once units dropped out, they could not return into the sample. Starting with $N = 2000$ units, nonresponse rate at each of the five time point hovers around .35.

As described in Section 4, error correlation in the panel selection model exists in the selection equation across time, in the substantive equation across time, and between selection and substantive equation at each time point. We generate two versions of simulations: The first version, the low-correlation version, has temporal variation in both the selection and the substantive equation ranging from .55 for adjacent time points to .45 for errors between time point one and time point five (Toeplitz structure). The correlation between selection and substantive equation at each time point is $\rho = -.2$. The second version, the high-correlation version, has temporal variation in both the selection and the substantive equation ranging from .85 to .56 (again, Toeplitz structure). The correlation between selection and substantive equation at each time point is $\rho = -.5$.

We estimate four models in each version of the Monte Carlo simulations. They compare the GEE estimator proposed in Section 4 under three different working correlation matrices with a GEE approach ignoring the selection based on unobserved variables. The last model simply ignores the selection equation and estimates the equation of interest, amounting to the estimation of a binary probit panel model. The structures of the three considered working correlation matrices in the GEE panel selection models are the identity matrix (GEE_I), equicorrelation (GEE_E), and Toeplitz structure (GEE_T). The same temporal error structure is assumed for the selection equation and the equation of interest. In the simple binary probit panel model in which the selection is ignored, we estimate a model based on a Toeplitz working correlation structure (GEE_T*).

Statistics calculated over 500 simulations under each condition are the mean (m) and SD of the estimates, the square root of the mean of estimated variances of the estimators, denoted as estimated standard deviation ($\widehat{\text{sd}}$), and the portion of rejections of the hypothesis H_0: $\beta_{js} = \beta_{js,0}$ (rej) for $\alpha = .05$, where s denotes the sth element of $\boldsymbol{\beta}_j$ and $\beta_{js,0}$ is the true value. The results are given in Table A1. Note that due to the lack of space, we only report simulation results for two regression parameters in both equations, that is, γ_1, γ_2, β_1, β_2, and for the error correlation between selection and substantive equation, ρ.

The results in Table A1 suggest that the estimators that account for the selection on the unobserved variables, GEE_I, GEE_E, and GEE_T, work quite well. That is, the mean of the estimates over the 500 simulations are close to the true values, the mean estimated standard errors are close to the standard deviations over the simulations, and the portions of rejections of the null are in an acceptable range of approximately $.05 \pm .02$.

Comparing the results for GEE_I versus GEE_E and GEE_T, we see that the gain in efficiency is mainly due to assuming dependence over time as compared to assuming independence. The gain in efficiency in adopting the equicorrelation assumption versus adopting the true Toeplitz structure is negligible.

Obviously, ignoring the selection on the unobserved variables (see GEE_T*) leads to severely biased estimators in the equation of interest and, thus, to rejection rates of 33.4% and 100% in the high-correlations condition and even 9.2% and 61% in the low-correlation condition.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Table A1 Mean (M), estimated standard deviation ($\widehat{sd}$), SD, and percentage of rejections of H_0: $\theta = \theta_0$ with $\alpha = .05$ (rej) over 500 simulations

	Low-correlation version				*High-correlation version*			
	GEE_I	GEE_E	GEE_T	GEE_T*	GEE_I	GEE_E	GEE_T	GEE_T*
				$\gamma_1 = .6$				
M	.601	.601	.601	.601	.600	.599	.599	.599
$\widehat{sd}$	.021	.020	.019	.019	.022	.020	.020	.020
SD	.021	.020	.020	.020	.022	.020	.020	.020
rej	.048	.054	.050	.050	.046	.068	.062	.062
				$\gamma_2 = -.7$				
M	−.701	−.701	−.701	−.701	−.701	−.700	−.700	−.700
$\widehat{sd}$	.032	.030	.029	.029	.032	.028	.028	.028
SD	.031	.029	.029	.029	.032	.027	.027	.027
rej	.066	.044	.042	.042	.050	.042	.048	.048
		$\rho = -.2$				$\rho = -.5$		
M	−.200	−.202	−.202	—	−.496	−.497	−.497	—
$\widehat{sd}$	.100	.101	.101	—	.080	.079	.079	—
SD	.097	.094	.093	—	.078	.074	.075	—
rej	.054	.044	.046	—	.050	.040	.046	—
				$\beta_1 = .5$				
M	.499	.499	.499	.555	.501	.501	.501	.659
$\widehat{SD}$	.039	.037	.037	.025	.041	.037	.037	.027
SD	.037	.035	.035	.024	.040	.036	.035	.026
rej	.044	.042	.042	.610	.044	.038	.036	1.000
				$\beta_2 = .3$				
M	.303	.303	.302	.273	.301	.300	.299	.232
$\widehat{sd}$	.050	.047	.047	.046	.047	.041	.041	.043
SD	.049	.046	.047	.046	.047	.040	.040	.043
rej	.056	.054	.052	.092	.050	.048	.044	.334

References

Abramson, Paul R., John H. Aldrich, and David W. Rohde. 1995. *Change and continuity in the 1992 elections*. Washington, DC: CQ Press.

Andrews, Donald W. 1998. Semiparametric estimation of a sample selection model. *Review of Economic Studies* 65:479–517.

Avery, Robert B., Lars P. Hansen, and V. Joseph Hotz. 1983. Multiperiod probit models and orthogonality condition estimation. *International Economic Review* 24:21–35.

Baker, Kendall L., Russel J. Dalton, and Kai Hildebrandt. 1981. *Germany transformed: Political culture and new politics*. Cambridge, MA: Harvard University Press.

Bartels, Larry M. 1994. The American public's defense spending preferences in the post-cold war era. *Public Opinion Quarterly* 58:479–508.

Beck, Paul A., and M. Kent Jennings. 1979. Political periods and political participation. *American Political Science Review* 73:737–50.

Berinsky, Adam J. 1999. The two faces of public opinion. *American Journal of Political Science* 43:1209–30.

Boehmke, Frederick J. 2003. Using auxiliary data to estimate selection bias models, with an application to interest group use of the direct initiative process. *Political Analysis* 11:234–54.

Bürklin, Wilhelm. 1987. Governing left parties frustrating the radical non-established left: The rise and inevitable decline of the Greens. *European Sociological Review* 4:161–6.

Bürklin, Wilhelm, and Russell J. Dalton. 1994. Das Ergrauen der Grünen. In *Wahlen und Wähler: Analysen aus Anlass der Bundestagswahl 1990*, eds. Hans-Dieter Klingemann and Max Kaase, 264–302. Pladen, Germany: Westdeutscher Verlag.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American voter*. New York: Wiley.
Dalton, Russell J., and Martin P. Wattenberg. eds. 2000. *Parties without partisans*. Oxford, UK: Oxford University Press.
Downs, Anthony. 1957. *An economic theory of democracy*. New York: Harper and Row.
Dubin, Jeffrey A., and Douglas Rivers. 1990. Selection bias in linear regression, logit and probit models. *Sociological Methods and Research* 18:360–90.
Dustmann, Christian, and María E. Rochina-Barrachina. 2007. Selection correction in panel data models: An application to the estimation of females' wage equations. *Econometrics Journal* 10:263–93.
Fiorina, Morris P. 2002. Parties and partisanship: A 40-year retrospective. *Political Behavior* 24:93–115.
Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. 1998. An analysis of sample attrition in panel data. *Journal of Human Ressources* 33:251–99.
Franklin, Mark N. 2004. *Voter turnout and the dynamics of electoral competition in established democracies since 1945*. Cambridge: Cambridge University Press.
Gayle, George-Levi, and Christelle Viauroux. 2007. Root-N consistent semiparametric estimators of a dynamic panel-sample-selection model. *Journal of Econometrics* 141:179–212.
Geddes, Barbara. 1990. How the cases you choose affect the answers you get: Selection bias in comparative politics. *Political Analysis* 2:131–50.
Goodliffe, Jay. 2001. The effect of war chests on challenger entry in U.S. House elections. *American Journal of Political Science* 45:830–44.
Green, Donald P., Bradley Palmquist, and Eric Schickler. 2002. *Partisan hearts and minds. Political parties and the social identities of voters*. New Haven, CT: Yale University Press.
Grier, Kevin B., Michael C. Munger, and Brian E. Roberts. 1994. The determinants of industry political activity, 1978–1986. *American Political Science Review* 88:911–26.
Hansen, Lars P. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50:1029–54.
Heckman, James J. 1974. Shadow prizes, market wages and labor supply. *Econometrica* 42:679–94.
———. 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economics and Social Measurement* 5:475–92.
———. 1979. Sample selection bias as a specific error. *Econometrica* 47:153–62.
Hug, Simon. 2000. Studying the electoral success of new political parties. *Party Politics* 6:187–97.
———. 2003. Selection bias in comparative research: The case of incomplete data sets. *Political Analysis* 11:255–74.
Hulsberg, Werner. 1988. *The German Greens: A social and political profile*. London: Verso.
Inglehart, Ronald. 1971. The silent revolution in Europe: Intergenerational change in post-industrial societies. *American Political Science Review* 65:991–1017.
Jackman, Simon, and Francis Vella. 1991. Redistricting and endogenous partisan control. *Political Analysis* 3:155–71.
Jennings, M. Kent. 1987. Residues of a movement: The aging of the American protest generation. *The American Political Science Review* 81:367–82.
Kitschelt, Herbert. 1988. The life expectancy of left-libertarian parties. Does structural transformation or economic decline explain party innovation? A response to Wilhelm Bürklin. *European Sociological Review* 4:155–60.
Klein, Markus, and Kai Arzheimer. 1997. Grau in Grau. Die Grünen und ihre Wähler nach eineinhalb Jahrzehnten. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 49:650–73.
Kohler, Ulrich. 1998. Zur Attraktivität der Grünen bei älteren Wählern. *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 50:536–59.
Kroh, Martin, and Peter Selb. 2009. Inheritance and the dynamics of partisanship. *Political Behavior* 31:559–74.
Kroh, Martin, and Martin Spiess. 2008. *Documentation of sample sizes and panel attrition in the German socio economic panel (SOEP) (1984 until 2007).* Number 39 in "Data documentation." Berlin, Germany: DIW.
Kyriazidou, Ekaterini. 1997. Estimation of a panel data sample selection model. *Econometrica* 65:1335–64.
Liang, Kung-Yee, and Scott L. Zeger. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22.
———. 1995. Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science* 10:158–73.
Little, Roderick J., and Donald B. Rubin. 2002. *Statistical analysis with missing data*. 2nd ed. New York: John Wiley.
Mannheim, Karl. 1928. Das Problem der Generationen. *Kölner Vierteljahreshefte für Soziologie* 7:157–85.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press

Manski, Charles F. 1989. Anatomy of the selection problem. *Journal of Human Resources* 24:343–60.
McCarty, Nolan, and Lawrence S. Rothenberg. 1996. Commitment and the campaign contribution contract. *American Journal of Political Science* 40:872–904.
Moore, Jeffrey C., Linda L. Stinson, and Edward J. Welniak. 1999. Income reporting in surveys: Cognitive issues and measurement error. In *Cognition and survey research*, eds. Monroe G. Sirken, Douglas J. Hermann, Susan Schechter, Norbert Schwarz, Judith M. Tanur, and Roger Tourangeau. 155–74. New York: Wiley.
Müller-Rommel, Ferdinand. 1985. The Greens in Western Europe: Similar but different. *International Political Science Review* 6:483–99.
Newey, Whitney K. 1999. Consistency of two-step sample selection estimators despite misspecification of distribution. *Economics Letters* 63:129–32.
Pannenberg, Markus, and Martin Spiess. 2009. GEE estimation of a two-equation panel data model with an application to wage dynamics and the incidence of profit-sharing in West Germany. *AStA Advances in Statistical Analysis* 93:427–47.
Richardson, Bradely M. 1991. European party loyalties revisited. *American Political Science Review* 85:751–75.
Rochina-Barrachina, M. E. 1999. A new estimator for panel data sample selection models. *Annales d'Économie et de Statistique* 55/56:153–81.
Rosenstone, Steven J., and John M. Hansen. 1993. *Mobilization, participation, and democracy.* New York: Macmillan Press.
Sartori, Ann E. 2003. An estimator for some binary-outcome selection models without exclusion restrictions. *Political Analysis* 11:111–38.
Schmitt-Beck, Rüdiger, Stefan Weick, and Bernhard Christoph. 2006. Shaky attachments: individual-level stability and change of partisanship among West German voters, 1984–2001. *European Jounrnal of Political Research* 45:581–608.
Teixeira, Ruy A. 1992. *The disappearing American voter.* Washington, DC: The Brookings Institution.
Timpone, Richard J. 1998. Structure, behavior, and voter turnout in the United States. *American Political Science Review* 92:145–58.
———. 2002. Estimating aggregate policy reform effects: New baselines for registration, participation, and representation. *Political Analysis* 10:154–77.
van der Eijk, Cees, Wouter van der Brug, Martin Kroh, and Mark N. Franklin. 2006. Rethinking the dependent variable in electoral behavior: On the measurement and analysis of utilities. *Electoral Studies* 25:423–46.
van der Klaauw, Bas, and Ruud H. Koning. 2003. Testing the normality Assumption in the sample selection model with an application to travel demand. *Journal of Business & Economic Statistics* 21(1):31–42.
Vella, Francis. 1998. Estimating models with sample selection bias: A survey. *Journal of Human Resources* 33:127–69.
Verba, Sidney, and Norman H. Nie. 1972. *Participation in America: Political democracy and social equality.* New York: Harper and Row.
Winship, Christopher, and Robert D. Mare. 1992. Models for sample selection bias. *Annual Review of Sociology* 18:327–50.
Wooldridge, Jeffrey M. 2002. *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.
Zorn, Christopher J. W. 2001. Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science* 45:470–90.
Zuckerman, Alan S. ed. 2005. *The social logic of partisanship*. Philadelphia, PA: Temple.
Zuckerman, Alan S., and Martin Kroh. 2006. The social logic of bounded partisanship in Germany: A comparison of West Germans, East Germans, and immigrants. *Comparative European Politics* 4:65–93.

https://doi.org/10.1093/pan/mpp045 Published online by Cambridge University Press