This is the Accepted Version. The Version of Record of this manuscript has been published and is available in Journal of Choice Modelling, Elsevier, 6 October 2015, https://doi.org/10.1016/j.jocm.2015.09.004
Mauerer, I., Pößnecker, W., Thurner, P. W., & Tutz, G. (2015). "Modeling Electoral Choices in Multiparty Systems with High-Dimensional Data: A Regularized Selection of Parameters Using the Lasso Approach." Journal of Choice Modelling 16: 23-42. doi: 10.1016/j.jocm.2015.09.004

Modeling electoral choices in multiparty systems with high-dimensional data: A regularized selection of parameters using the Lasso approach

Ingrid Mauerer[1], Wolfgang Pößnecker[2], Paul W. Thurner[1], and Gerhard Tutz[2]

[1]Geschwister-Scholl-Institute for Political Research, University of Munich (LMU)
[2]Department of Statistics, University of Munich (LMU)

September, 2015

Accepted: Journal of Choice Modelling

Abstract

The increased usage of discrete choice models in the analysis of multiparty elections faces one severe challenge: the proliferation of parameters, resulting in high-dimensional and difficult-to-interpret models. E.g., the application of a multinomial logit model in a party system with J parties results in maximally $J-1$ parameters for chooser-specific attributes (e.g., sex, age). For the specification of alternative-specific attributes (usually: positions on issues and issue distances), maximally J parameters for each political issue can be estimated. Thus, a model of party choice with five parties based on three political issues and ten voter attributes already produces 59 possible coefficients. As soon as we allow for interaction effects to detect segment-specific reactions to issues, the situation is even aggravated. In order to systematically and efficiently identify relevant predictors in voting models, we derive and use Lasso-type regularized parameter selection techniques that take into account both individual- and alternative-specific variables. Most importantly, our new algorithm can handle for the first time the alternative-wise specification of the attributes of alternatives. Applying the specifically adjusted Lasso method to the 2009 German Parliamentary Election, we demonstrate that our approach massively reduces the models' complexity and simplifies their interpretation. Lasso-penalization clearly outperforms the simple ML estimator. The results are illustrated by innovative visualization methods, so-called effect star plots.

Keywords: Parameter Selection, Lasso, Multinomial Logit Model, Multiparty Elections.

1 Introduction

The usage of discrete choice models in the analysis of voters' choice behavior in multiparty systems has become common practice. Multinomial logit and probit models[1] are well-established tools in electoral research. Extensive analyses based on voter characteristics (age, sex, income etc.) on the one hand and attributes of the parties or candidates on the other hand[2] contributed enormously to our understanding of the strategies of party competition (such as position-taking and issue saliency) and of voters' individual choice behavior (see Alvarez and Nagler, 1998; Adams, Merrill, and Grofman, 2005). However, the well-known flexibility of discrete choice models comes along with a severe drawback: the proliferation of possible coefficients. This is the reason why we propose for the first time to consider sophisticated parameter selection techniques and to adjust them to this specific context of discrete choice modeling. To illustrate the problem more precisely, party choice in a party system with J parties results in maximally $J-1$ parameters in the case of chooser-specific attributes. For each alternative-specific attribute, maximally J parameters can be estimated. As a consequence, the amount of possible individual- and alternative-specific coefficients increases rapidly, resulting in highly complex and difficult-to-interpret models. Moreover, as soon as we allow for interaction effects, e.g., in order to test for segment-specific reactions to issue distances, the situation is even aggravated. For these reasons, the following research question is raised: How can we systematically and efficiently reduce this high-dimensional parameter space?

In this paper, we address the problem of parameter proliferation in discrete choice modeling. Our objective is to efficiently model (electoral) choices based on very high-dimensional data. We introduce and derive a Lasso-type regularization technique in the estimation of multinomial logit models (MNLs) which takes into account both individual- and alternative-specific variables. Most importantly, our new tailor made solution allows us to handle for the first time the alternative-wise specification of choice-specific covariates[3] (e.g., issue distances).[4]

[1]In the following we use the general term "multinomial logit model" to refer to multinomial logit models including both alternative- and individual-specific variables. In addition, we use the terms "alternative-specific", "choice-specific" and "party-specific" covariates interchangeably to refer to attributes of alternatives (i.e., parties/candidates). Analogously, "chooser-specific", "voter-specific" or "individual-specific" variables refer to voter characteristics.

[2]E.g., the distance between voter's most preferred position on scales of controversial issues and the corresponding position of these alternatives.

[3]In contrast to Tutz, Pößnecker, and Uhlmann (2015), who propose the so-called Categorically Structured Lasso (CATS Lasso) to penalize groups formed by all coefficients that belong to the same individual-specific variable in MNLs, we introduce for the first time a tailor made adjustment of the Lasso that explicitly incorporates the alternative-wise specification of choice-specific covariates, and offer a in-depth discussion of the resulting variable selection and its practical implications in the context of party choice modeling.

[4]While the alternative-wise specification of choice-specific attributes is widely used and applied in transportation economics and econometrics (see, e.g., Ben-Akiva and Lerman, 1985; Hensher, Rose, and Greene, 2007; Louviere, Hensher, and Swait, 2009; Train, 2009), it has received little attention in the empirical study of voting behavior. See as an early exception Thurner (2000), and the more detailed recent illustration by Mauerer, Thurner, and Debus (2015).

The Lasso approach is a regularized parameter selection technique that penalizes the L_1-norm of the coefficients. This enforces parameter selection and reduction of the predictor space, and therefore guarantees – in contrast to classical subset selection approaches – continuous, stable and computationally efficient variable selection.

By applying the Lasso method to the 2009 German Parliamentary Election, we demonstrate that our proposed approach massively reduces the model's complexity and simplifies its interpretation by selecting the most important determinants of party choice. In addition, we show that Lasso-penalization improves the model's predictive performance. More specifically, we demonstrate that Lasso-penalization clearly outperforms the simple ML estimator, for both a main effects and an interaction effects model. The results are summarized by using so-called effect star plots.

The contribution of this paper is threefold. First, it introduces and illustrates the benefits of Lasso-type regularization techniques in the study of choice behavior, and therefore offers a highly promising solution to the problem of high-dimensional predictor spaces in MNLs. Second, in contrast to previous work, our new Lasso method not only takes into account both individual- and alternative-specific variables, but also is able to handle the alternative-wise specification of choice-specific variables. Even though the benefits of our Lasso approach are illustrated by using electoral choices, it can be applied to any choice situation based on this type of model class facing the problem of parameter inflation.[5] Third, we show that using a symmetric identifiability constraint prevents that the arbitrary choice of a reference category biases the Lasso's selection of the most promising model.

The remainder of the paper is structured as follows: To introduce the problem of high-dimensional data in modeling electoral choices in multiparty systems, Section 2 provides a short formal outline of the theoretical model. Then, we briefly examine classical variable selection procedures and their limitations. In particular, we demonstrate why and how the usage of regularization methods in the study of multiparty elections enables us to efficiently identify important predictors, and therefore to improve electoral choice models, and to facilitate their interpretation. In Section 4 we illustrate the usefulness of the proposed approach by providing a regularized analysis of party choice in the 2009 German Parliamentary Election. Finally, Section 5 summarizes the major advantages of our approach in the analysis of electoral choice behavior.

2 Model Formulation

MNLs are mainly used in the study of individual voting behavior to translate the Spatial Theory of Voting into a statistical model.[6] The Spatial Theory of Voting assumes that

[5] All codes and routines to implement the proposed model will be made available on the authors' website.

[6] Since Downs' seminal work on democracy as a political market (Downs, 1957), a growing number of scholars theoretically as well as technically developed this approach (see, e.g., Alvarez and Nagler, 1998;

voter i chooses the party/candidate j that offers policy positions on K policy dimensions, denoted by p_{ijk}, that are closest to the voter's most preferred policy position on each k dimension, denoted by x_{ik}. According to the principle of utility maximization, voter i compares the parties' policy proposals p_{ijk}[7] and identifies each party's supplied amount of utility, denoted by $V_{ij}, j = 1, ..., J$:

$$V_{ij} = -\left(\sum_{k=1}^{K} \alpha_k \left|x_{ik} - p_{ijk}\right|\right), \tag{1}$$

where α_k presents the weight or saliency of each kth policy dimension.[8] These coefficients are the utility parameters indicating the value attached to the dimensions. The observed part of utility V_{ij} consists of two components that are connected additively: alternative-specific and individual-specific variables. Alternative-specific variables $z_{ijk},\ k = 1, \ldots, K$, represent the distance between voter i and party j on policy dimension k and is defined by $z_{ijk} = -|x_{ik} - p_{ijk}|$. Individual-specific variables $s_{il},\ l = 1, \ldots, p$, refer to voter characteristics.[9] Thus, for $i = 1, \ldots, n$ and $j = 1, \ldots, J$, the deterministic part of utility takes the following form:[10]

$$\eta_{ij} = \beta_{j0} + \sum_{l=1}^{p} s_{il}\beta_{jl} + \sum_{k=1}^{K} z_{ijk}\alpha_{jk} = \beta_{j0} + \boldsymbol{s}_i^T \boldsymbol{\beta}_j + \boldsymbol{z}_{ij}^T \boldsymbol{\alpha}_j.$$

The parameters $\beta_{10}, \ldots, \beta_{J0}$ represent alternative-specific constants (ASCs). Each $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J$ is a p-dimensional coefficient vector related to the p-dimensional individual-specific covariate vector $\boldsymbol{s}_i$. The coefficient vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_J$ contain the issue saliency parameters that indicate the value attached to the K policies/issues in the alternative-specific covariate vector $\boldsymbol{z}_{ij}$. Note the used parameterization of the effect of alternative-specific variables: In contrast to the restrictive set-up with generic coefficients in (1) (i.e., estimating one coefficient for all alternatives), we apply the alternative-wise specification which allows estimating J different parameters α_{jk} in the case of J alternatives, so that different effects on alternatives are possible. The alternative-wise specification of choice-specific variables – which is widely known and applied in the discrete choice literature

Thurner, 2000; Dow and Endersby, 2004; Adams, Merrill, and Grofman, 2005; Schofield et al., 1998). For a summary of the formal set-up of the Spatial Theory of Voting, see Thurner (2000).

[7]Note that we explicitly use respondent-specific perceptions of party positions instead of perceived mean party placements. By using subjective self-placements of party positions, we follow previous studies who offer both theoretical arguments and empirical evidence in favor of voter-specific perceptions of party positions (see, e.g., Gilljam, 1997; Lewis and King, 1999; Westholm, 1997; Merrill III and Grofman, 1997).

[8]For a legitimation of the disaggregate City-Block metric in the calculation of issue distances, see Singh (2014).

[9]Note here that classical spatial voting models present policy-only models, i.e., they are exclusively based on the alternative-specific variables "issue distances". By incorporating voter-specific variables, we follow the postulate of Adams, Merrill, and Grofman (2005) who especially emphasize the important role of voters' non-policy or non-spatial characteristics/considerations in models of issue voting.

[10]V_{ij} and η_{ij}, the linear predictor, denote the same quantity, accumulating the observable determinants of the vote decision process in a scalar quantity. Since "linear predictor" is the common denomination in the statistical literature, we prefer this term in discussions about formal model-setup.

(see, e.g., Ben-Akiva and Lerman, 1985; Train, 2009; Hensher, Rose, and Greene, 2007; Louviere, Hensher, and Swait, 2009) – has received little attention in the empirical study of voting behavior so far. By estimating party-specific issue saliency parameters, our model specification is very much in line with recent developments in the conception of political parties' competition on issues. Several studies have shown that issue competition is characterized by both position-taking on issues and by differently attaching saliencies to issues (see, e.g., Green-Pedersen, 2007; Guinaudeau and Persico, 2014; Meguid, 2005; Rovny, 2012; Wagner, 2012). Therefore, parties induce issue voting not only by taking positions but also by attaching more or less saliency to issues, resulting in voters' issue saliencies that vary across both parties and issues. In a recent study, Mauerer, Thurner, and Debus (2015) have shown that these two aspects of issue competition can be systematically combined by estimating party-varying reactions to issues. We explicitly follow these new developments and insights by estimating party-specific issue saliency parameters; i.e., we explicitly relax and test the assumption that voters attach the same importance to each party's position along a given dimension.

Considering both individual-specific predictors and the alternative-wise specification of choice-specific variables, the multinomial logit model (MNL) in its generic form can be stated as follows (see Tutz, 2012):[11]

$$\pi_{ij} = P(Y = j|\boldsymbol{s}_i, \boldsymbol{z}_{ij}) = \frac{\exp(\eta_{ij})}{\sum_{r=1}^{J} \exp(\eta_{ir})} = \frac{\exp(\beta_{j0} + \boldsymbol{s}_i^T \boldsymbol{\beta}_j + \boldsymbol{z}_{ij}^T \boldsymbol{\alpha}_j)}{\sum_{r=1}^{J} \exp(\beta_{r0} + \boldsymbol{s}_i^T \boldsymbol{\beta}_r + \boldsymbol{z}_{ij}^T \boldsymbol{\alpha}_r)}, \tag{2}$$

where $Y \in \{1, ..., J\}$ denotes the j-categorical, probabilistic response variable, i.e. $Y = j$ indicates that party j is chosen.[12] Note that (2) refers to the MNL in its generic form, which means that the parameters $\beta_{10}, ..., \beta_{J0}$ and $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_J$ are not identifiable. In order to identify the model, a side constraint, such as defining a reference category or using a symmetric side constraint, has to be introduced.[13]

At this point it is important to emphasize the inherent complexity of the MNL in (2). As the following example indicates, the number of possible individual- and party-specific coefficients increases rapidly, resulting in highly complex and difficult-to-interpret models. Consider a model of party choice in a system with five major parties based on three choice-specific variables and ten chooser-specific attributes. This specification already produces 59 possible parameters (ASCs included).[14] The following section introduces the

[11] Logit models assume that the random part of utility, ϵ_{ij}, follows an iid maximum extreme value distribution. See also McFadden (1974).

[12] For later use, we define J-dimensional response vectors $\boldsymbol{y}_i = (0, ..., 0, 1, 0, ..., 0)^T$ with 1 on jth position indicating the chosen alternative (i.e. $Y = j$). Additionally, let $\boldsymbol{\pi}_i$ denote the J-dimensional vector of choice probabilities. Conditional on the covariates $\boldsymbol{s}_i$ and $\boldsymbol{z}_{ij}$, $\boldsymbol{y}_i$ can be considered as independent realizations of drawing from a multinomial distribution: $\boldsymbol{y}_i \mid \boldsymbol{s}_i, \boldsymbol{z}_{ij} \sim M(1, \boldsymbol{\pi}_i),\ i = 1, \dots, n$.

[13] However, in Section 3.2.3, we will show an interdependency between the particular choice of an identifiability constraint and the Lasso method. Therefore, the discussion of identifiability is delayed to Section 3.2.3.

[14] If a reference category as side constraint is used, we obtain 4 ASCs + 4*10 coefficients for individual-specific variables + 3*5 parameters for party-specific variables, resulting in a total of 59 parameters.

Lasso method, a parameter selection strategy that systematically and efficiently reduces this high-dimensional predictor space, and thus ensures more parsimonious spatial voting models.

3 The Lasso Approach: Parameter Selection and Regularization of MNLs with alternative-specific Covariates

The previous section highlighted the inherent complexity of MNLs, and therefore the practical need for sophisticated parameter selection procedures in the analysis of electoral choice behavior based on the Spatial Theory of Voting. In this section, we introduce and derive for the first time a Lasso-type regularization technique for the estimation of MNLs that considers both individual- and alternative-specific variables. Additionally, this new algorithm allows us to handle the alternative-wise specification of the attributes of alternatives. Based on a brief review of classical subset selection procedures and their weaknesses and limitations, we outline the general idea of regularization and penalty approaches. Then, the Lasso approach is presented, including a technical discussion of its computation, its interdependency with the choice of identifiability constraint in MNLs, and how its variable selection properties can be improved.

3.1 Classical Subset Selection Techniques

How can electoral researchers systematically select the most promising model out of a large set of possible models (based on a large number of potential predictor variables) while simultaneously reducing the model's complexity? Since choosing a model that maximizes some goodness-of-fit measure (e.g., pseudo R^2 or loglikelihood) causes overfitting and low predictive accuracy, the task of variable selection is typically tackled by introducing optimality criteria that approximate a given model's expected performance on future observations. Two of the most popular optimality criteria are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).[15] Classical subset selection techniques, such as best-subset[16] or stepwise selection techniques[17], investigate

In the case of a symmetric side constraint, 70 nominal parameters result, while still obtaining only 59 degrees of freedom. See Section 3.2.3.

[15] For the specified MNL considering the alternative-wise specification of issue distances, these optimality criteria are given by

$$\text{AIC}(\hat{\boldsymbol{\theta}}) = -2l(\hat{\boldsymbol{\theta}}) + 2\text{df}(\hat{\boldsymbol{\theta}}); \qquad \text{BIC} = -2l(\hat{\boldsymbol{\theta}}) + \log(n)\text{df}(\hat{\boldsymbol{\theta}}), \tag{3}$$

where $\hat{\boldsymbol{\theta}}^T = (\hat{\beta}_{10}, \ldots, \hat{\beta}_{J0}, \hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_J, \hat{\boldsymbol{\alpha}}_1, \ldots, \hat{\boldsymbol{\alpha}}_J)$ denotes the estimator of the model's overall parameter vector $\boldsymbol{\theta}$, $l(\boldsymbol{\theta})$ the loglikelihood, and $\text{df}(\hat{\boldsymbol{\theta}})$ the degrees of freedom, equalling here the number of parameters.

[16] Best-subset refers to the choice of the, in terms of optimality criteria, best possible subset of variables out of all possible variable combinations.

[17] Stepwise selection approaches include forward selection, backward selection, or a combination thereof. Stepwise approaches can be seen as an attempt to approximate best-subset selection with lower compu-

the influence of the inclusion or exclusion of individual predictors on minimizing these optimality criteria. As a result, the combination of covariates yielding the smallest value of the optimality criteria is chosen.

However, these frequently used classical subset selection approaches exhibit several weaknesses. With regard to best-subset selection procedures, it is rarely possible to compute exactly the optimal set of variables due to the associated computational burden. Take, for example, a model of party choice consisting of five parties, three issues and ten chooser-specific attributes (see Section 4). Applying the best-subset selection technique to this model, 55 parameters could be set to zero[18], resulting in a total of $2^{55} \approx 3.6 \times 10^{16}$ possible models. This example demonstrates a side-effect of the flexibility of the MNL framework: Since the number of parameters is the product of the number of alternatives and predictors, it is impracticable to fit all possible models, unless the number of alternatives and predictors is extremely small.

In order to obtain a satisfying subset of important predictors in reasonable time, one typically applies stepwise approaches in which variables are added or removed from the current state/model until the optimality criteria cannot be improved any more. However, since subset selection is a discrete process and all optimality criteria have multiple local optima, these stepwise approaches suffer from considerable instability (Hastie, Tibshirani, and Friedman, 2009). Thus, starting stepwise variable selection either from a full model or starting it from an ASCs-only model can lead to completely different results. Consequently, even the slightest change in the data or in the starting point can produce significantly different outcomes of subset selection. Due to this instability, stepwise approaches cannot be recommended, but also an exhaustive all-subset search is usually impossible for the considered model class. In the next section, we present the Lasso approach. In contrast to classical subset selection techniques, the Lasso is a parameter selection method guaranteeing continuous, stable and computationally efficient variable selection.

3.2 The Lasso

In order to motivate the Lasso method that efficiently identifies relevant predictors, we briefly outline the general idea of regularization techniques, of which the Lasso is a special case. Regularization implies, inter alia, introducing penalty terms that restrict the estimated coefficients (see Tutz, 2010). Therefore, penalization refers to the formulation of side constraints on the values of the parameters which are taken into account in the estimation. Penalty approaches based on L_p-norm of the parameter vector aim to penalize the size or the length of these parameters. Their goal is to shrink the coefficients and to set, ideally, coefficients of weak predictors exactly to zero, yielding continuous, stable

tational burden. For an overview of variable selection based on subset techniques, see Hastie, Tibshirani, and Friedman (2009) and references therein.

[18]This number of parameters implies that ASCs always remain in the model.

and computationally efficient variable selection. In the following, we first introduce the Lasso approach, including its definition and basic properties. Second, we outline how the Lasso's variable selection properties can be improved by using adaptive weights. Then, we demonstrate the Lasso's interdependency with the choice of identifiability constraint in MNLs, followed by a technical discussion on the computation of the Lasso estimator and the choice of the tuning parameters λ.

3.2.1 Definition and Basic Properties

The Lasso (Least Absolute Shrinkage and Selection Operator), introduced by Tibshirani (1996), is a penalty approach to variable selection in regression models. For a general model with loglikelihood $l(\cdot)$ and parameter vector $\boldsymbol{\theta}$[19], penalty approaches introduce a penalty term $P(\boldsymbol{\theta})$ that is subtracted from the loglikelihood, resulting in the penalized loglikelihood $l_{\text{pen}}(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - P(\boldsymbol{\theta})$. That is, instead of maximizing the loglikelihood in the estimation process, we maximize the penalized loglikelihood. If the penalty term is chosen appropriately, the penalized parameter estimator can have superior properties compared to the unpenalized maximum likelihood (ML) estimator, such as a reduced variance or a lower dimensionality. The p-norm penalized parameter estimator is, in its most general form, defined by

$$\hat{\boldsymbol{\theta}}^{pen}(\lambda, p) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ l(\boldsymbol{\theta}) - \lambda||\boldsymbol{\theta}||_p = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ - l(\boldsymbol{\theta}) + \lambda||\boldsymbol{\theta}||_p, \qquad p \geq 0, \tag{4}$$

where $\lambda \geq 0$ denotes the tuning parameter controlling the degree of penalization. The Lasso estimator is obtained by penalizing the L_1-norm of the parameter vector: $P_{\text{Lasso}}(\boldsymbol{\theta}) = \lambda||\boldsymbol{\theta}||_1 = \lambda \sum_i |\theta_i|$.[20]

In the following, we explicitly derive for the first time the Lasso for the MNL based on the predictor structure given in (2) and discuss its properties. The Lasso has previously been extended to MNLs by Friedman, Hastie, and Tibshirani (2010). However, their work mainly focuses on algorithms and only briefly mentions the MNL as a possible application for their algorithm. Additionally, Friedman, Hastie, and Tibshirani (2010) exclusively consider individual-specific predictors $\boldsymbol{s}_i$, whereas we explicitly derive the Lasso for a MNL containing both individual- and choice-specific predictors. Therefore, and in contrast to previous work, which only briefly mentions the MNL as a possible application and exclusively considers individual-specific predictors, we explicitly derive the Lasso for a MNL based on both individual- and alternative-specific predictors and in which the alternative-specific variables are specified as alternative-specific effects.

[19]For the specified model including issue distances with alternative-specific effects, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$, where $\boldsymbol{\beta} = (\beta_{10}, \dots, \beta_{J0}, \boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_J)$ is the vector of all β-parameters and, accordingly, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_J)$.

[20]Penalizing the squared L_2-norm results in ridge regression (Hoerl and Kennard, 1970). Note that best subset selection based on AIC or BIC is also contained in (4) by using the L_0-pseudo-norm and particular values of λ.

Applying the general form of the Lasso to our model yields[21]

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmax}}\; l(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \lambda \sum_{j=1}^{J} \left(\sum_{l=1}^{p} |\beta_{jl}| + \sum_{k=1}^{K} |\alpha_{jk}| \right). \tag{5}$$

The Lasso solutions may contain exact zeros, so that the corresponding effects are effectively removed from the model. Thus, the Lasso implicitly performs variable selection as a by-product of the estimation process. The larger λ, the stronger is the penalization, and therefore the sparser is the estimated coefficient vector. Setting $\lambda = 0$ leads to the ML estimator. A maximal value $\lambda_{\max}$ can be derived so that all penalized parameters are estimated to be zero whenever $\lambda \geq \lambda_{\max}$. Since the Lasso estimator is the maximizer of a continuous and concave objective, it is continuous itself, unique for any fixed λ[22] and less sensitive to noise in the data than the discrete best subset method. In fact, the choice $p = 1$ in (4), which defines the Lasso, is the only choice for which the L_p-norm penalized log-likelihood is concave and, at the same time, yields a sparse estimator; i.e., is able to contain exact zeros.[23] Consequently, the Lasso can perform variable selection just like subset selection, but avoids its drawbacks by being continuous in the data and computationally efficient.

More insights into the Lasso can be gained by considering the following alternative representation. Based on standard results from optimization theory (cf. Boyd and Vandenberghe, 2004), the Lasso can equivalently be defined by a constrained optimization:[24]

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmax}}\; l(\boldsymbol{\beta}, \boldsymbol{\alpha}) \qquad \text{subject to} \quad \sum_{j=1}^{J} \left(\sum_{l=1}^{p} |\beta_{jl}| + \sum_{k=1}^{K} |\alpha_{jk}| \right) \leq t. \tag{6}$$

The tuning parameter $t \geq 0$ of the Lasso's constrained definition is connected to λ in its penalized form by a one-to-one mapping, but deriving a closed form for their relationship is difficult. By defining $t_{\max} = \sum_{j=1}^{J} \left(\sum_{l=1}^{p} |\hat{\beta}_{jl}^{ML}| + \sum_{k=1}^{K} |\hat{\alpha}_{jk}^{ML}| \right)$, the Lasso estimator is equal to the ML estimator for any $t \geq t_{\max}$. For $t < t_{\max}$, the constraint in (6) induces shrinkage of the parameters. As this constrained definition of the Lasso is highly useful to illustrate why the Lasso is able to shrink coefficients to exactly zero, consider the geometric illustration in Figure 1. The figure shows the contour lines of the log-likelihood function

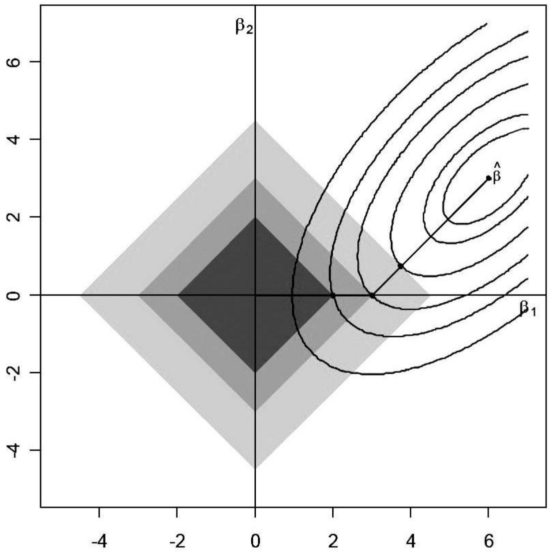
[21] Note that if alternative J is chosen as reference to achieve identifiability of the MNL, we set $\beta_{J0} = 0$, $\boldsymbol{\beta}_J = \mathbf{0}$. Hence, summing up over all J alternatives in (5) is equivalent to penalizing only $J-1$ β-vectors that actually have to be estimated. Therefore, formula (5) is applicable to all MNLs, regardless of the chosen identifiability constraint. Also note that we do not penalize the ASCs. To keep the notation readable, we omit the dependence of Lasso estimates on a particular choice of the tuning parameter λ.

[22] The objective function is concave, and thus has a unique maximum. Assuming a sensible design, i.e. no avoidable multicollinearity, uniqueness of the parameter vector that attains this maximum is ensured if $n > J(1 + p + K)$. This is virtually always the case for applications in political and social sciences.

[23] Formally, this follows from two facts: 1.) The L_p-norm is nonconvex for $p < 1$, leading to local maxima of the penalized log-likelihood, and thus to a discrete variable selection method. 2.) For $p > 1$, the L_p-norm is differentiable at zero, and hence cannot yield solutions containing exact zeros.

[24] More specifically, equivalence between the constrained and penalized definition of the Lasso follows from the so-called Karush-Kuhn-Tucker conditions, which extend the well-known Langrange multiplier method to optimization under inequality constraints.

A.



Note: Lasso estimates are given by the points of contact between log-likelihood contour lines and Lasso constraint regions.

Figure 1: Lasso geometry: constraint regions for the Lasso with log-likelihood contour lines for two predictors in a simple logistic regression model.

and the Lasso constraint regions for different values of t in a two-dimensional predictor space; i.e., based on a simple logistic regression model with two predictors and without an intercept. The Lasso solutions must necessarily lie at the contact points of the contour lines with the constraint regions, indicated by the black line. The ML estimator is at the center of the log-likelihood contour lines. As the inspection of Figure 1 demonstrates, the diamond-shape of the Lasso constraint causes the corresponding Lasso solutions to move in a straight line towards the axis when t is successively reduced, eventually shrinking $\hat{\beta}_2$ to zero.

The tendency of the Lasso to produce sparse solutions can also be demonstrated by considering the special case of a linear model with orthonormal design: Let $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. By assuming that $\boldsymbol{y}$ and all predictors $\boldsymbol{x}_l$ $(l = 1, \ldots, p)$ are mean-centered, no intercept is used, and that $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}$, the Lasso has an analytical solution: $\hat{\beta}_l^{Lasso} = \text{sign}(\hat{\beta}_l^{ML}) \cdot \max\left(|\hat{\beta}_l^{ML}| - \lambda, 0\right), \ l = 1, \ldots, p$. Hence, the lasso solution for each predictor in this particular setting is obtained by shrinking the respective ML estimator towards zero by a value of λ. Since the MNL is a nonlinear model, even an orthonormal design matrix does not allow us to derive a similar result (or any analytical solution), but the functioning of the Lasso in MNLs is very similar. Additionally to yielding parsimonious solutions and due to the fact that the Lasso is a shrinkage estimator, the Lasso estimates have reduced size and variability. These aspects, as well as the computation of the effective degrees of freedom are discussed and summarized in Appendix A.

3.2.2 Improving the Lasso using adaptive Weights

In an $n \to \infty$ setting (i.e., with arbitrarily large sample size), a variable selection method should guarantee the selection of the correct model, that is to assign nonzero estimates to truly nonzero effects and to set the coefficients of irrelevant predictors to zero. In addition, it is desirable that the method asymptotically performs as well as the ML estimator, applied to the correct set of variables. If an estimator possesses these two properties, it is said to be as good as an 'oracle' that knows the correct set of variables ahead of time. As shown by Zou (2006), the ordinary Lasso from (5) does not possess this oracle property because it applies the same degree of penalization / shrinkage to all parameters, regardless of their specific size. Choosing λ as large as it is necessary to remove all irrelevant predictors implies too much bias on the estimated coefficients of the selected variables. How can we avoid that the same degree of penalization is applied to all parameters, regardless of their specific size, and therefore improve the Lasso solutions? The remedy proposed by Zou (2006) is the so-called adaptive Lasso, which is defined as follows:

$$
\begin{gathered}
(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmax}}\, l(\boldsymbol{\beta}, \boldsymbol{\alpha}) - \lambda \sum_{j=1}^{J} \left(\sum_{l=1}^{p} w^b_{jl}\, |\beta_{jl}| + \sum_{k=1}^{K} w^a_{jk}\, |\alpha_{jk}| \right), \\
w^b_{jl} = \frac{1}{|\hat{\beta}^{ML}_{jl}|}, \quad w^a_{jk} = \frac{1}{|\hat{\alpha}^{ML}_{jk}|}, \qquad l = 1, \dots, p;\; k = 1, \dots, K;\; j = 1, \dots, J.
\end{gathered} \tag{7}
$$

As formula (7) shows, the adaptive Lasso incorporates an individual weight for the penalty of each coefficient. Note that each weight (w^b_{jl} and w^a_{jk}) consists of the inverse of the size of the corresponding ML estimate. By using these weights, the adaptive Lasso is able to adapt more adequately to the varying importance of different variables for different alternatives / parties. Since the ML estimator is consistent, the ML estimates of irrelevant effects asymptotically turn to zero, yielding an infinite (or at least very large) penalization. In the case of nonzero effects, the adaptive weights converge to a finite, constant value. In general, the larger the ML estimate of a particular parameter, the less penalization is applied to this parameter by the adaptive Lasso, and vice versa. This logic also holds in non-asymptotic settings as long as the ML estimator is stable. Indeed, our experience has shown that the use of adaptive weights provides a considerable improvement of the Lasso's performance, both in terms of variable selection and prediction accuracy – even for data sets with a moderate number of observations. For these reasons and even though the adaptive Lasso was originally motivated by theoretical, asymptotic considerations, we strongly recommend its usage in virtually all practical applications.

3.2.3 Lasso and its Interdependency with the Choice of Identifiability Constraint

As we noted in Section 2, not all individual-specific variables (and ASCs) in (2), giving the MNL in its generic form, are identifiable. In order to identify the model, a side constraint has to be introduced. In this section we point out how the choice of the identifiability constraint in MNLs and the result of the Lasso interdepend. In particular, we show that the use of a symmetric identifiability constraint prevents that the arbitrary choice of a reference category influences the results of the Lasso method.

The non-identifiability of all individual-specific covariates in MNLs can be expressed in the following way: for $l = 0, 1, \dots, p$, let $\delta_l \in \mathbb{R}$ denote a shift which is applied to all coefficients belonging to the same individual-specific predictor, so that the new coefficients $\tilde{\beta}_{jl} = \beta_{jl} + \delta_l$ result. By gathering this shift across all p individual-specific predictors into a shift vector $\boldsymbol{\delta}^T = (\delta_1, \dots, \delta_p)$, the model based on $\tilde{\beta}$-parameters, for $j = 1, \dots, J$, can be expressed as follows:[25]

$$\begin{aligned} P(Y = j|\boldsymbol{s}, \boldsymbol{z}) &= \frac{\exp(\tilde{\beta}_{j0} + \boldsymbol{s}^T\tilde{\boldsymbol{\beta}}_j + \boldsymbol{z}_j^T\boldsymbol{\alpha}_j)}{\sum_{r=1}^{J}\exp(\tilde{\beta}_{r0} + \boldsymbol{s}^T\tilde{\boldsymbol{\beta}}_r + \boldsymbol{z}_r^T\boldsymbol{\alpha}_r)} \\ &= \frac{\exp(\beta_{j0} + \delta_0 + \boldsymbol{s}^T(\boldsymbol{\beta}_j + \boldsymbol{\delta}) + \boldsymbol{z}_j^T\boldsymbol{\alpha}_j)}{\sum_{r=1}^{J}\exp(\beta_{r0} + \delta_0 + \boldsymbol{s}^T(\boldsymbol{\beta}_r + \boldsymbol{\delta}) + \boldsymbol{z}_r^T\boldsymbol{\alpha}_r)} \\ &= \frac{\exp(\delta_0 + \boldsymbol{s}^T\boldsymbol{\delta}) \cdot \exp(\beta_{j0} + \boldsymbol{s}^T\boldsymbol{\beta}_j + \boldsymbol{z}_j^T\boldsymbol{\alpha}_j)}{\exp(\delta_0 + \boldsymbol{s}^T\boldsymbol{\delta}) \cdot \sum_{r=1}^{J}\exp(\beta_{r0} + \boldsymbol{s}^T\boldsymbol{\beta}_r + \boldsymbol{z}_r^T\boldsymbol{\alpha}_r)}. \end{aligned} \tag{8}$$

Equation (8) shows that the absolute level of the β-parameters is not relevant for the MNL, only the differences $\beta_{jl} - \beta_{ql},\ j \neq q,\ l = 0, 1, \dots, p$ are meaningful. Since the Lasso penalty is based on this absolute value of the coefficients, the choice of the identifiability constraint in MNLs influences the results of the Lasso method, and therefore the selection of the most promising model.

In general, scholars choose as identifiability constraint a particular reference category. By using $J = 1$ as reference/baseline category, β_{10} and $\boldsymbol{\beta}_1$ are set to zero. This ensures identifiability, and the effects of individual-specific covariates are interpreted relative to the first alternative/party. In order to highlight the interplay between the choice of a particular reference category as identifiability constraint and the Lasso, consider an unpenalized model with five parties, in which the first party is chosen as reference. Assume that the ML-coefficients of variable s_l for this model are given by $\beta_l = (0, 3, 0.2, -0.5, 2.5)^T$, so that $||\beta_l||_1 = 6.2$. By defining the second party as reference instead, it follows from (8) that β_l changes to $(-3, 0, -2.8, -3.5, 0.5)$, and thus $||\beta_l||_1 = 9.8$. As this simple example demonstrates, the arbitrary change of the reference category causes to increase both the

[25]Note that this kind of parameter shifting does not change the model.

overall L_1-norm and the number of coefficients that are relatively far away from zero. Consequently, the sparsity of the Lasso solutions also decreases – at least for some values of λ. Using real data, we illustrate in Section 4.1 that the sparsity and predictive performance of the Lasso indeed vary considerably across different reference categories.

Hence, how can we prevent that the arbitrary choice of a reference category influences the results of the Lasso method? We propose to use a symmetric identifiability constraint, which results from imposing that

$$\sum_{j=1}^{J} \beta_{jl} = 0, \qquad l = 0, 1, \dots, p. \tag{9}$$

Using this symmetric side constraint yields J ASCs and $J{*}p$ parameters in the case of individual-specific predictors. However, only $J-1$ and $(J-1){*}p$ of these parameters can be estimated freely. As Zou, Hastie, and Tibshirani (2007) showed, the effective degrees of freedom of the Lasso estimator are generically given by the number of nonzero parameters. By combining this general template with the symmetric side constraint (i.e., for the estimator from (5) coupled with (9)), the following formula for the Lasso's effective degrees of freedom results:[26]

$$\hat{\mathrm{df}}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \sum_{l=0}^{p} \max\left(\Big(\sum_{j=1}^{J} I(|\hat{\beta}_{jl}| > 0)\Big) - 1,\ 0\right) \ + \sum_{k=1}^{K}\sum_{j=1}^{J} I(|\hat{\alpha}_{jk}| > 0). \tag{10}$$

Finally, in the next section we discuss the computation of the Lasso estimator and the choice of the tuning parameter λ.

3.2.4 Estimation and Choice of Tuning Parameters

In order to compute the Lasso estimator for any given λ, we use the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) introduced by Beck and Teboulle (2009). After adjusting the formulas for the loglikelihood and its derivatives, the FISTA algorithm can be directly applied to our Lasso specification ((5) or (7)). For the specified MNL including issue distances with alternative-specific effects, the loglikelihood is computed as

$$l(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij} \log\big(\pi_{ij}(\boldsymbol{\beta}, \boldsymbol{\alpha})\big), \tag{11}$$

where π_{ij} is given in formula (2) and is written here as a function of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. For $j = 1, \dots, J$, the derivatives of the loglikelihood with respect to the parameters are

[26] $I(\cdot)$ denotes an indicator function.

obtained as follows:

$$\begin{aligned}\frac{\partial l(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \beta_{jl}} &= \sum_{i=1}^{n} s_{il}(y_{ij}-\pi_{ij}) = \boldsymbol{s}_l^T(\boldsymbol{y}_j-\boldsymbol{\pi}_j), \quad l=0,1,\ldots,p,\\ \frac{\partial l(\boldsymbol{\beta},\boldsymbol{\alpha})}{\partial \alpha_{jk}} &= \sum_{i=1}^{n} z_{ijk}(y_{ij}-\pi_{ij}) = \boldsymbol{z}_{jk}^T(\boldsymbol{y}_j-\boldsymbol{\pi}_j), \quad k=1,\ldots,K,\end{aligned} \tag{12}$$

where $\boldsymbol{s}_l^T = (s_{1l},\ldots,s_{nl})$, $\boldsymbol{z}_{jk}^T = (z_{1jk},\ldots,z_{njk})$, $\boldsymbol{y}_j^T = (y_{1j},\ldots,y_{nj})$, and $\boldsymbol{\pi}_j^T = (\pi_{1j},\ldots,\pi_{nj})$ denote vectors pooling the corresponding quantities across n observations. To sum up, we offer the Lasso specification as well as the computation of the loglikelihood and its derivatives for a MNL including alternative-specific variables with alternative-specific effects.

Since the Lasso penalty is not invariant to the scale and the variance of the covariate vectors $\boldsymbol{s}_l$ and $\boldsymbol{z}_{jk}$, we ensure that all variables have the identical chance of being selected by standardizing these to zero mean and unit variance before applying the Lasso. This leads to Lasso-penalized regression coefficients for the standardized covariates from which the corresponding coefficients belonging to the original covariates can easily be reconstructed.

In the following, we outline the choice of the tuning parameter λ, the parameter controlling the degree of penalization. To choose an appropriate λ, the Lasso solutions are computed over a grid of different λ-values. By m denoting the number of possible λs, this grid is given by a sequence $\lambda_1 > \lambda_2 > \ldots > \lambda_m$. For practical reasons, the lower endpoint λ_m of this grid usually consists of a very small, but positive value (e.g., $\lambda_m = 0.01$.) The corresponding upper endpoint is chosen as $\lambda_1 = \lambda_{\max}$, which is the smallest value of the tuning parameter for which all penalized coefficients are set to zero. Using standard results from convex optimization theory, it is easy to show that $\lambda_{\max}$ can be computed as

$$\lambda_{\max} = \max\left(\max_{j,l}\left|\boldsymbol{s}_l^T\left(\boldsymbol{y}_j-\hat{\boldsymbol{\pi}}_j^0\right)\right|,\ \max_{j,k}\left|\boldsymbol{z}_{jk}^T\left(\boldsymbol{y}_j-\hat{\boldsymbol{\pi}}_j^0\right)\right|\right),$$

where $\hat{\boldsymbol{\pi}}_j^0$ presents the estimated choice probabilities of a model containing only unpenalized predictors. Note that in our case this is an ASCs-only model.

In order to select a concrete λ and thus a concrete estimator / model, the Lasso solutions for different values of the tuning parameter are evaluated by model selection criteria. Usually, these optimality criteria are either crossvalidation (CV), AIC or BIC (which can be computed applying (10)). According to our experience, the selection of λ based on BIC leads to the most sparse and CV to the least sparse model. Since the AIC presents a desirable compromise, we strongly prefer to use this optimality criterion for the choice of λ. Here it is important to note that variable selection using the Lasso solely requires a one-dimensional search over a grid of λ-values, whereas the complexity of variable selection based on best subset techniques or test-based procedures increases exponentially with the number of covariates and alternatives.

4 Application: Regularized Analysis of Party Choice in the 2009 German Parliamentary Election

This section illustrates the advantages of the Lasso-type regularization method in the analysis of multiparty competition using individual survey data from the 2009 German Parliamentary Election.[27] The survey contains information on 2173 respondents. Our data set includes 816 cases (and $816 \cdot 5 = 4080$ observations/rows) resulting from casewise-deletion. That is, only those respondents are included that (1) placed themselves and all the parties on each issue, and (2) have no missing values on all individual-specific variables, and (3) reported voting for one of the five major parties. Party choice is operationalized via the stated vote intention. The set of alternatives includes parties having a vote share total of at least 5% in the 2009 election. The respective parties are: the Christian-Democratic Party (CDU), the Social-Democratic Party (SPD), the Liberal Party (FPD), the Green Party (Greens), and the Leftist Party (Leftists). In order to demonstrate the Lasso's potential to effectively and systematically reduce the predictor space to important determinants of individual vote choices, our model of party choice is based on the following explanatory variables: As alternative-specific predictors we include three position issues and the Left-Right ideological scale, measured as the absolute distance between the individually perceived party positions and the self-reported positions of voters (ideal points) on 11-point scales.[28] As individual-specific covariates we consider 10 voter characteristics that are widely used to explain party choice in Germany. These voter characteristics are sex, age, West/East Germany, union membership, high school degree, unemployment, political interest, satisfaction with democracy, and religious denomination.[29] These voter attributes represent important non-policy factors and reflect central social cleavage structures influencing party choice in Germany.

The presentation of the practical usefulness of the regularized analysis of party choice is divided into two parts: We begin by fitting a Lasso-penalized MNL based on solely

[27] German Longitudinal Election Study (GLES) 2009, Pre-election Cross-Section. The data set is available under http://www.gesis.org/en/elections-home/gles/data-and-documents/data/. Identification Number: ZA5300, Version 5.0.0.

[28] The position issues are: Taxes: "1" = Lower taxes, even if that means less government spending on health, education and social benefits, "11" = More government spending on health, education and social benefits, even if that means higher taxes; Immigration: "1" = Laws on immigration should be relaxed, "11" = Laws on immigration should be tougher; Nuclear Energy: "1" = More nuclear power stations, "11" = Close down all nuclear power stations immediately. Note again that we use voter-specific perceptions of party positions (see Section 2).

[29] These variables are coded as follows: sex: 1 (male), 0 (female); age: centered around the sample mean of 50.5 years, measured in decades, metric; West/East Germany: 1 (former West Germany), 0 (former East Germany); union membership: 1 (union members) , 0 (otherwise); high school degree: 1 (yes), 0 (no); unemployment: 1 (currently unemployed), 0 (otherwise); political interest: 1 (less interested), 0 (very interested); satisfaction with democracy: 1 (not satisfied), 0 (satisfied); religious denomination: (Protestant, Roman-Catholic, otherwise).

main effects. As different segments may react differently to issue distances, we also test for segment-specific reactions to issues. Therefore, in the second part we additionally include interactions between issues and voter attributes. In each section, we provide a systematic comparison of the penalized and unpenalized models.

4.1 Main Effects Model

As a first step of our analysis, we estimated a Lasso-penalized MNL of party choice containing the main effects of all variables introduced above. As argued in Section 3.2.3, we applied a symmetric side constraint to ensure identifiability of the model and to avoid that the arbitrary choice of a reference category influences the results of the Lasso method.[30] We fit this model over a grid of 100 λ-values and chose the best one according to the AIC-criterion. Following the discussion from Section 3.2.2, adaptive weights are included in the Lasso penalty. Since there exists no analytical formula for standard errors of Lasso estimates, we computed them via bootstrap.[31] In order to illustrate the difference between an unpenalized MNL of party model and a Lasso-penalized MNL, we contrast the results of these models. The parameter estimates of the unpenalized ML main effects model are given in Table 1. Table 2 summarizes the parameter estimates as well as approximate p-values (based on the bootstrapped standard errors) of the corresponding Lasso-regularized model.[32]

As outlined in Section 2, we apply the alternative-wise specification to choice-specific covariates. In order to find the most suitable specification for the effect of the alternative-specific variables, we examined the equality of the resulting parameters using Likelihood-Ratio (LR) tests. The results indicate that only for the issue of taxes one fixed generic coefficient is sufficient, whereas the issue of immigration and nuclear energy require alternative-specific coefficients. That is, with regard to these two issues, the party-specific issue saliency coefficients statistically differ from each other, indicating that splitting up the generic coefficient is necessary. To model the effect of the ideological Left-Right scale, the generic specification is appropriate. Therefore, the effects of the Left-Right scale and the issue of taxes are fixed across parties, and the issues of immigration and nuclear energy are equipped with party-specific issue saliency coefficients (see at the bottom of Table 1 and 2.)[33]

As can be immediately seen from Table 2, a considerable amount of effects is set to zero by the Lasso, yielding an enormous reduction of the complexity of the estimated model. In particular, 31 out of 67 nominal parameters are set to zero, and thus only 36 coefficients are selected; i.e., remain in the model, and therefore only this subset of effects

[30] More details on how the Lasso's sparsity and predictive performance vary across different reference categories will be given below.

[31] Our bootstrapped standard error estimates are based on $B = 500$ bootstrap samples.

[32] Note that the p-values of coefficients which are set to zero by the Lasso are necessarily 1. Therefore, we do not display these p-values.

[33] The LR test between the chosen model and a model where all issues have generic effects yields a p-value of 0.018.

	CDU		SPD		FDP		Greens		Leftist	
ASC	1.32***	(0.000)	0.06	(0.860)	-0.53	(0.202)	-0.41	(0.413)	-0.44	(0.289)
Sex	-0.20	(0.237)	-0.26	(0.103)	0.05	(0.827)	0.08	(0.717)	0.33	(0.124)
West Germany	-0.57**	(0.003)	0.53*	(0.016)	-0.27	(0.304)	0.50	(0.087)	-0.18	(0.474)
Age	0.02***	(0.000)	0.01	(0.082)	-0.01	(0.542)	-0.02**	(0.003)	0.00	(0.981)
Union Membership	-0.52	(0.084)	0.39	(0.085)	-0.33	(0.308)	-0.48	(0.156)	0.95***	(0.000)
High School	0.25	(0.279)	-0.12	(0.594)	0.27	(0.341)	0.34	(0.187)	-0.74*	(0.022)
Unemployment	-0.04	(0.933)	0.05	(0.912)	-0.15	(0.867)	-0.22	(0.720)	0.36	(0.450)
Pol. Interest	0.17	(0.369)	0.27	(0.166)	-0.27	(0.248)	0.09	(0.740)	-0.25	(0.301)
Democracy	-0.69***	(0.000)	-0.29	(0.105)	0.17	(0.443)	-0.17	(0.459)	0.99***	(0.000)
ReligionCatholic	0.57**	(0.009)	-0.20	(0.301)	0.33	(0.227)	-0.16	(0.571)	-0.53*	(0.049)
ReligionOther	-0.08	(0.692)	-0.29	(0.151)	0.56*	(0.038)	0.05	(0.848)	-0.23	(0.398)
Taxes	0.16***	(0.000)	0.16***	(0.000)	0.16***	(0.000)	0.16***	(0.000)	0.16***	(0.000)
Immigration	0.20**	(0.001)	0.04	(0.441)	0.06	(0.399)	0.16**	(0.006)	0.09	(0.190)
Nuclear Energy	0.12*	(0.025)	0.22**	(0.001)	0.23**	(0.007)	0.25***	(0.000)	0.12	(0.124)
Left-Right	0.63***	(0.000)	0.63***	(0.000)	0.63***	(0.000)	0.63***	(0.000)	0.63***	(0.000)

Numbers in parentheses show approximate p-values based on simple, two-sided t-tests using bootstrapped standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: ML coefficient estimates of the main effects model, 2009 German Parliamentary Election.

	CDU		SPD		FDP		Greens		Leftist	
ASC	1.30***	(0.000)	-0.11	(0.602)	-0.43	(0.127)	-0.09	(0.539)	-0.67	(0.104)
Sex	0		-0.17	(0.212)	0		0		0.17	(0.140)
West Germany	-0.53*	(0.015)	0.39*	(0.041)	0		0.14	(0.382)	0	
Age	0.02**	(0.001)	0		0		-0.02***	(0.000)	0	
Union Membership	-0.61	(0.059)	0		0		0		0.61*	(0.011)
High School	0		0		0		0.36	(0.069)	-0.36	(0.123)
Unemployment	0		0		0		0		0	
Pol. Interest	0		0.11	(0.248)	-0.11	(0.320)	0		0	
Democracy	-0.74***	(0.000)	0		0		0		0.74***	(0.000)
ReligionCatholic	0.10	(0.092)	0		-0.10	(0.612)	0		0	
ReligionOther	-0.26	(0.416)	0		0.26	(0.096)	0		0	
Taxes	0.16***	(0.000)	0.16***	(0.000)	0.16***	(0.000)	0.16***	(0.000)	0.16***	(0.000)
Immigration	0.18**	(0.001)	0		0		0.15**	(0.005)	0.06	(0.257)
Nuclear Energy	0.11*	(0.025)	0.22***	(0.001)	0.23**	(0.006)	0.25***	(0.000)	0.10	(0.151)
Left-Right	0.62***	(0.000)	0.62***	(0.000)	0.62***	(0.000)	0.62***	(0.000)	0.62***	(0.000)

Numbers in parentheses show approximate p-values based on simple, two-sided t-tests using bootstrapped standard errors. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Lasso-regularized coefficient estimates of the main effects model, 2009 German Parliamentary Election.

substantially influence party choice.[34]

We accentuate the following points resulting from Table 2. First, inspecting the issue effects[35] at the bottom of Table 1 and 2 shows that (a) the fixed generic effects of

[34] These 67 nominal parameters result from the symmetric side constraint and the generic specification of the issue of taxes and the Left-Right scale, but only 56 degrees of freedom are obtained, of which 24 degrees of freedom are actually selected by the Lasso.

[35] Also note that all issue effects are positive. Therefore, since the issue distance covariates z_{ijk} are defined as $z_{ijk} := -|x_{ik} - p_{ijk}|$, a larger difference between the individually perceived party positions and the voter's ideal points on these issues indicates that this party is less likely to be chosen by this voter. That is, if the perceived distance increases by one unit, the corresponding issue variable decreases by one unit due to the negative sign.

the issue of taxes and the Left-Right scale are selected by the Lasso. As expected, the Left-Right scale exhibits the largest effect on party choice. (b) Even though all party-specific coefficients for the issue of nuclear energy remain in the model, the parameters substantially vary across parties and not all of them are significant. (c) For the issue of immigration only three out of five possible party-specific saliency parameters are selected by the Lasso, but only the CDU and the Greens trigger significant issue reactions. Only these two parties – who offer clear, non-ambivalent polar positions on the issue of immigration (CDU: contra immigration; Greens: pro immigration) – seem to successfully attract voters by their position-taking and saliency efforts. The results clearly indicate that issue considerations strongly determine vote choice. In most cases, the impact of issues on party choice substantially varies across parties, implying that not all parties are equally successful in attracting voters based on their issue-related campaign efforts. In contrast to the unpenalized MNL, the Lasso-regularized MNL allows us to systematically remove the non-substantive party-specific parameters from the party choice model.

Second, one voter attribute – unemployment status – is entirely removed from the model, suggesting that this variable does not influence party choice at all.

Third, only subsets of the remaining individual-specific effects are selected by the Lasso, indicating that only specific voter attributes are decisive for choosing particular parties. Notice, e.g., the highly significant age effect on the vote for the Christian-Democratic Party CDU and the Greens. As to be expected, the CDU is strong among older voters, whereas the Greens seem to be attractive among younger voters. In particular, by employing the Lasso-based selection technique, we can show that age has only an impact on voting for the CDU and the Greens; i.e., there is no age effect with regard to the remaining parties.

A series of further interesting effects has been selected by the Lasso, including the strength of the Leftist Party among voters being members of a labor union, or the positive effect of having a Catholic affiliation on supporting the CDU. These two findings are particularly relevant for researchers focusing on the impact of traditional cleavages on party choice. In line with recent work (see, e.g., Elff, 2009, 2007), these selected effects support their claim that social and class divisions continue to be relevant determinants of electoral behavior in Germany. However, note that only the effect of the variable union membership on the Leftist vote is selected and highly significant, whereas the one for the Social-Democratic Party SPD does not remain in the model. This result may indicate that the SPD has become less effective in attracting and mobilizing its traditional target and that the class cleavage is more effectively represented by the Leftist Party. The reason why the social cleavage seems to find less expression in the SPD might be seen as the result of unpopular labor market reforms (Hartz IV) initiated by Chancellor Schroeder in 1998. Again, in contrast to the unpenalized model, where all effects of the covariate union membership are included – with remarkable size –, the Lasso-regularized model contains only those that are relevant for party choice, thereby considerably facilitating the interpretation of the impact of union membership on party choice.[36]

Also note that our results suggest that the persistent relevance of traditional cleavages does not as strongly apply to religious versus secular divisions. Although the Lasso selected a positive effect of having a Catholic affiliation on supporting the CDU, and a negative effect on the FDP vote, they do not prove to be significant, indicating an increasing secularisation of the German society.[37] Since Christian affiliation constituted for a long time a major cleavage in the German party system, this is an important result. Finally, also note the interesting and highly significant selected effect of voter's satisfaction with democracy. Voters who are unsatisfied with the democracy in Germany are much less likely to vote for the CDU, while they are much more likely to vote for the Leftists. Thus, out of the five parties considered here, the Leftist Party seems to be the strongest attractor for 'protest voters'.

On this basis, it can be concluded that the shrinkage and regularization effect of the Lasso systematically and efficiently removes unimportant non-substantive parameters from the party choice model. Consequently, only relevant determinants of individual vote choices remain in the model. This greatly reduces the number of parameters to be interpreted. To be precise, the unselected coefficients do not have to be considered in the prediction of aggregate vote shares/market shares under different hypothetical scenarios, e.g., to evaluate how party movements on the issue scales change choice probabilities. These unselected, unimportant terms can be simply ignored. With regard to this model class and the applied model specification this presents a huge practical advantage because we have to set fewer variables to specific values when predicting central quantities of interest. In addition, the Lasso clearly outperforms the ML estimator with regard to performance measures. To demonstrate this, we compare the Lasso-penalized main effects model with the corresponding unpenalized model. The models' predictive performance is measured by the cross-validated deviance (CV), the AIC, and the BIC. The models' complexity is determined by the effective degrees of freedom.[38] The results of this comparison are presented in Table 3. As Table 3 shows, the Lasso considerably improves all four considered predictive performance and complexity measures compared to the ML estimator.

The coefficients from Table 2 can be interpreted as usual. However, their interpreta-

[36]Also note the following: Even though the coefficients of the covariate union membership have the same absolute value for both the CDU and the Leftist Party (due to the symmetric side constraint), the effect on the Leftist vote is significant while the effect on the CDU vote is not. This is no contradiction because the variance of the effect of a dummy-coded binary predictor on a particular party depends on the percentage of observations having chosen this party, and the corresponding predictor simultaneously took the value coded with 1. Consequently, these two estimated effects can differ in variance. In general, it should be noted that the p-values and significance levels shown in Table 2 are based on simple t-tests, and thus ignore the correlation between predictors. Therefore, a selected parameter not gaining significance according to this test should not cause concern.

[37]However, note here that we operationalize the religious-secular cleavage by religious denomination without additionally considering the frequency of church attendance.

[38]For an unpenalized MNL with symmetric side constraint, effective degrees of freedom equal the number of nominal parameters minus the number of individual-specific variables.

Model	CV	AIC	BIC	$\widehat{\text{df}}$
ML main effects	180.58	1697.55	1960.99	56
Lasso main effects	171.38	1656.98	1798.40	24

Table 3: Comparison of main effects models based on predictive performance and complexity.

tion is enormously simplified by considering only important substantive coefficients. For example, for voters from former West Germany relative to those living in former East Germany, the odds of voting for the CDU relative to voting for the SPD change by a factor of $\exp(-0.53 - 0.39) = 0.399$, if everything else remains fixed. Thus, the odds of voting for CDU instead of SPD are roughly 60% lower in former West Germany, ceteris paribus. In contrast to the individual-specific variables and as the following example demonstrates, the alternative-specific issue distance coefficients have to be interpreted slightly differently. If the perceived distance between a voter and the Green Party on the issue of immigration increases by x units, the odds of voting for the Green Party relative to voting for any other party change by a factor of $\exp(-0.15 \cdot x) = 0.86^x$, given the other variables in the model are held constant.[39] Hence, ceteris paribus, a one-unit increase in the distance on the issue of immigration decreases the odds of voting for the Greens by 14%, a two-unit increase decreases the odds by 26% etc.[40]

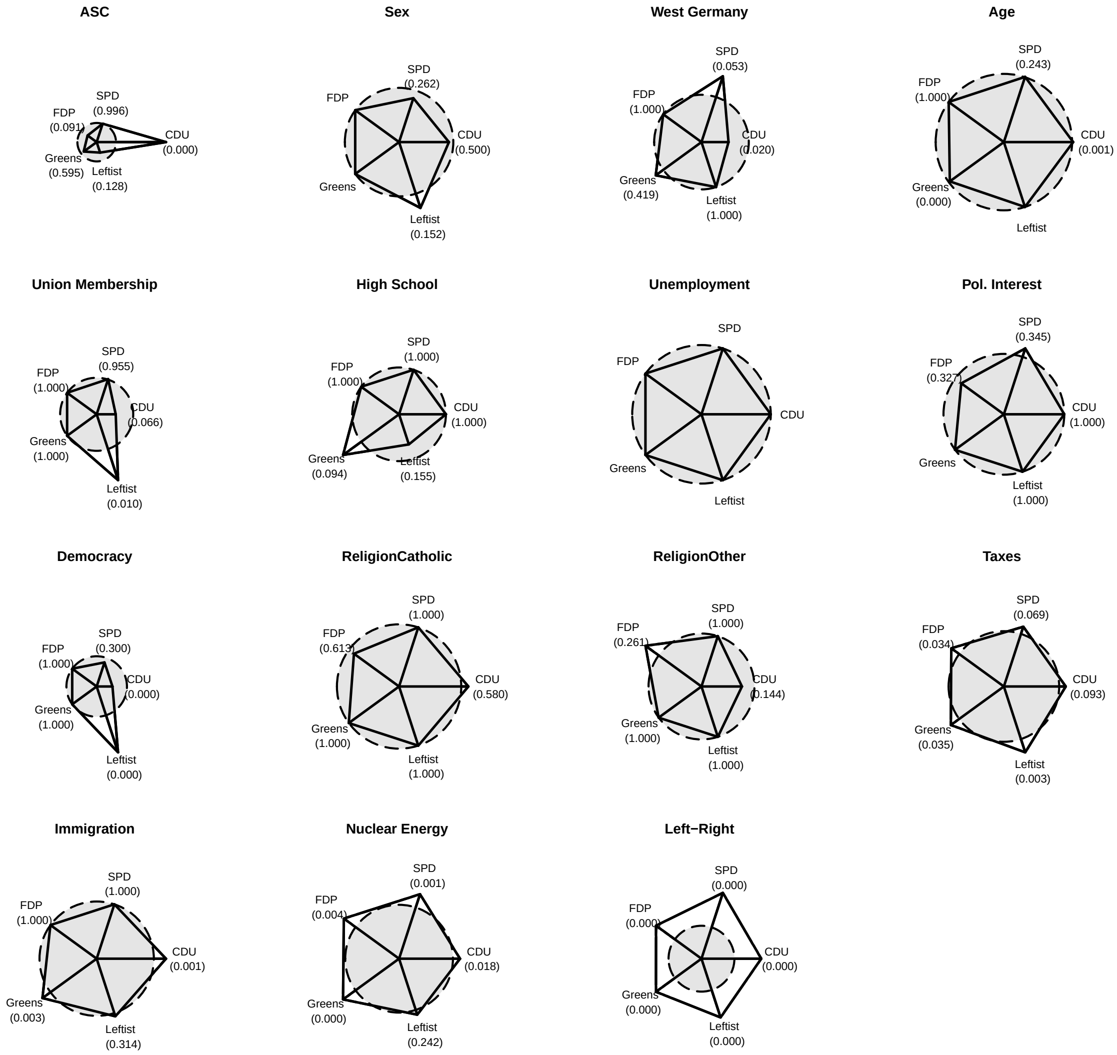
To summarize the main effects model in Table 2 in a concise and easy-to-read fashion, we suggest the use of so-called effect stars (Tutz and Schauberger, 2013). Figure 2 depicts these effect stars for each covariate. Each ray of a particular effect star corresponds to one party and has length proportional to $\exp(\hat{\beta}_{jl})$ or $\exp(\hat{\alpha}_{jk})$, respectively. The dashed circles correspond to a null effect, and therefore indicate a length of $\exp(0) = 1$. If the ray of an effect star lies outside of this circle, the corresponding variable exhibits a positive effect on the respective party; if the ray lies inside, the effect is negative. The numbers in parentheses show approximate p-values.[41] Since the overall size of all effect stars is the same, small circles correspond to predictors whose maximal effect is large, whereas large circles indicate a small maximal effect. For instance, visual inspection shows that the covariate unemployment does not affect party choice, all rays of its effect star lie on the null circle. Also note the striking large effect of dissatisfaction with democracy on the Leftist Party, or the strong negative effect of union membership on the CDU vote.

Next, let us turn to the visualization of the Lasso's shrinkage property as outlined in Section 3.2.1. This shrinkage effect can be illustrated by so-called coefficient paths. Figure 3 depicts these coefficient paths for the variable Political Interest and the issue of immigration. Each path indicates the Lasso estimates over the chosen grid of the

[39] Note again that due to the negative sign of the issue distances, an increase in the perceived distance corresponds to a decrease in the issue variable.

[40] $\exp(-0.15 \cdot 2) = 0.74$; $1 - 0.74 = 0.26 = 26\%$.

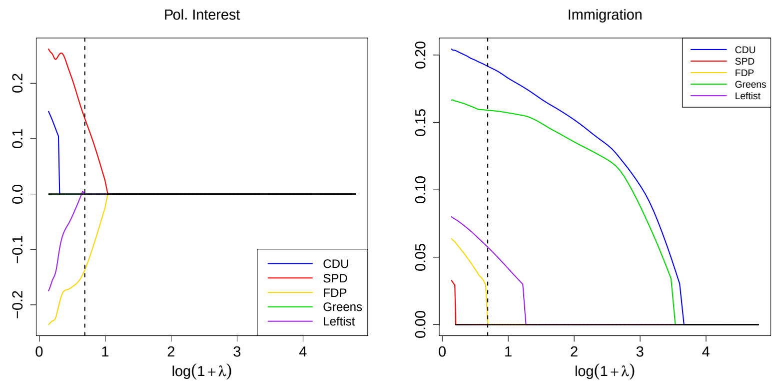
[41] Note again that the p-values of coefficients which are set to zero by the Lasso are necessarily 1. Therefore, we do not display these p-values. For instance, see the effect of sex on the Greens vote.



Note: The effect stars are based on the main effects model in Table 2.

Figure 2: Effect stars, 2009 German Parliamentary Election.

tuning parameter λ. In particular, the paths illustrate how the estimated coefficients move towards zero when the penalization is increased. Therefore, they show for which specific party these two variables have the most persistent effect. With regard to the issue of immigration, these parties are the Leftist, the Greens and the CDU: Only by applying a strong penalization – particularly on the parameters of the Greens and the CDU – the issue coefficients turn to zero. The horizontal black line indicates a zero effect, the dashed vertical line visualizes the λ chosen via AIC. Hence, the height of the intersection between coefficient paths and the dashed line is equal to the estimators from Table 2. Take, for instance, the path of the SPD for the issue of immigration. This path intersects the dashed line at about .2, visualizing the Lasso-regularized coefficient estimate for the SPD vote. In order to facilitate the readability of the coefficient paths, we use a log-scale



Note: The coefficient paths are based on the main effects model in Table 2.

Figure 3: Coefficient paths for the variables Political Interest and the issue of immigration, 2009 German Parliamentary Election.

for the x-axis. By applying the transformation "log(1 + λ)", the paths begin at $\lambda = 0$, which corresponds to the ML estimator. Note that the left panel of Figure 3 does not contain a green path representing the Green Party. This means that the effect of the variable political interest on the Greens vote is already shrunk to zero by the Lasso even for $\lambda_{min} = 0.15$, presenting the smallest degree of penalization in our grid. The complete presentation of coefficient paths can be found in Figure 4 in Appendix B.

Finally, we show how the parsimony and predictive performance of the Lasso vary across different reference categories. To support our arguments from Section 3.2.3, Table 4 summarizes the performance measures for the Lasso-penalized main effects model based on the six possible identifiability contraints (i.e., five choices of reference category and symmetric side constraint). It can be immediately seen that the sparsity of the

Identifiability Contraint	CV	AIC	BIC	$\widehat{\text{df}}$
CDU	168.61	1674.82	1860.20	31
SPD	168.73	1674.73	1869.62	32
FDP	167.91	1671.19	1843.56	28
Greens	168.20	1668.33	1845.07	27
Leftist	169.38	1674.76	1873.47	32
Symmetric	171.38	1656.98	1798.40	24

Table 4: Comparison of performance measures for Lasso-penalized main effects model based on different choices of the identifiability constraint.

models varies substantially across different choices of the reference category, indicated by the amount of nonzero parameters. Therefore, the Lasso's predictive performance also varies considerably across reference categories. Applying the symmetric side constraint provides in this application the most parsimonious model and also performs best in terms of AIC and BIC. This observed pattern, however, cannot be generalized: the relative performance of a model applying a symmetric side constraint compared to one using a reference category depends on the data set at hand. Nonetheless, the symmetric side constraint provides a general solution to the identifiability issue in MNLs while the choice of a specific reference category is always arbitrary.

4.2 Including Interactions between Issues and Voter Attributes

As it is obvious that different segments may react differently to issue distances, a natural strategy is to systematically test for segment-specific reactions to issues. Therefore, we consider a model allowing for interaction effects between issues and voter attributes. In particular, by using this much more complex specification we are able to determine the issue effects for specific voter demographics. Note that each issue distance parameter of the main effects model estimated in the previous section shows the effect of an issue on a party for **all** kinds of voters; i.e., the marginal issue effect across the whole population. Although the alternative-specific specification of the issue variables already allows us to detect which particular issue proves to affect which particular party choice, the main effects model does not allow us to infer which specific voter segments place a differential emphasis on the issues when casting their ballots. By segmenting the population into subgroups, we are also able to identify so-called issue publics (Converse, 1964). According to the issue public hypothesis, the population can be divided into issue publics, each consisting of voters who intensively care about particular issues. Instead of assuming that voters are homogeneously sensitive to the whole spectrum of issues, the issue public hypothesis suggests that specific voter segments can be distinguished by their differing sensitivities towards issues based on, e.g., their personal interests or demographic characteristics (see, e.g., Krosnick, 1990; Thurner, 2000; Mebane, Jackson, and Wall, 2014). The Lasso-based regularization technique allows us to systematically detect those issue publics. While the use of interactions is attractive from both a theoretical and practical perspective, it massively increases the model's complexity and aggravates its interpretation. For instance, our application case consists of three issue variables, ten voter attributes and five parties, resulting in a total of 150 possible interaction terms. This high-dimensional interaction model cannot be properly handled by unpenalized ML estimation. In particular, fitting a model of this size without penalization yields highly unstable estimates with poor predictive performance. Thus, the main challenge is the continuous, stable and computationally efficient selection of substantive segment-specific reactions to issues out of a large set of possible interaction terms.

Using the same methodology and steps as in the previous section, we fit a Lasso-penalized model that additionally includes interactions between the issues and voter characteristics. Motivated by previous research, we pre-selected the following segment-specific reactions to issues: $issue_{kj} \cdot education_i$, $issue_{kj} \cdot political\ interest_i$, $issue\ of\ immigration_j \cdot religion_i$. The first two sets of interactions reflect the political sophistication hypothesis, stating that the higher the level of political sophistication (as measured by political interest and level of education), the larger the impact of issues on party choice (see, e.g., Luskin, 1987). Since the interaction model is also based on the alternative-wise specification of the issue saliency parameters, it allows us to detect whether the moderating effect of political sophistication on issue voting equally applies to all parties and issues. The third set of interactions examines whether religious denomination causes voters to attach different saliencies to the issue of immigration when they evaluate the parties on this issue. The specified interaction model allows us to infer whether the three considered voter segments differ in the importance they attach to issues when they evaluate each of the parties. Table 5 contains the Lasso-penalized parameter estimates of this interaction model.[42]

	CDU		**SPD**		**FDP**		**Greens**		**Leftist**	
ASC	1.60***	(0.000)	-0.16	(0.550)	-0.58	(0.102)	-0.18	(0.612)	-0.67*	(0.035)
Sex	0		-0.17	(0.265)	0		0		0.17	(0.208)
West Germany	-0.59**	(0.009)	0.37	(0.077)	0		0.23	(0.310)	0	
Age	0.02**	(0.002)	0		0		-0.02***	(0.001)	0	
Union Membership	-0.48	(0.163)	0.16	(0.459)	0		-0.46	(0.183)	0.79**	(0.002)
High School	0		0		0		0		0	
Unemployment	0		0		0		0		0	
Pol. Interest	0		0		0		0		0	
Democracy	-0.75***	(0.000)	0		-0.07	(0.950)	0		0.82***	(0.000)
ReligionCatholic	-0.32	(0.297)	0		0.89*	(0.018)	0		-0.57	(0.094)
ReligionOther	-0.70*	(0.035)	0		1.15**	(0.003)	0		-0.45	(0.182)
Taxes	0.16**	(0.004)	0.16**	(0.004)	0.16**	(0.004)	0.16**	(0.004)	0.16**	(0.004)
Immigration	0.36*	(0.011)	0		0.07	(0.533)	0.21*	(0.016)	0	
Nuclear Energy	0.22**	(0.006)	0.25*	(0.035)	0.24**	(0.008)	0.24**	(0.001)	0.10	(0.187)
Left-Right	0.63***	(0.000)	0.63***	(0.000)	0.63***	(0.000)	0.63***	(0.000)	0.63***	(0.000)
Taxes X High School	0		0		0		0		0	
Taxes X Pol. Interest	0		0		0		0		0	
Immigration X High School	-0.11	(0.245)	0		-0.12	(0.347)	-0.15	(0.184)	0	
Immigration X Pol. Interest	0.16	(0.121)	0		-0.27	(0.060)	0		0.17	(0.052)
Immigration X ReligionCatholic	-0.38**	(0.007)	-0.04	(0.595)	0.30	(0.098)	0		0.02	(0.742)
Immigration X ReligionOther	-0.28*	(0.047)	0.09	(0.302)	0.24	(0.101)	0		-0.05	(0.568)
Nuclear Energy X High School	0		0		0		0		0	
Nuclear Energy X Pol. Interest	-0.14	(0.094)	-0.05	(0.624)	0		0		0	

Numbers in parentheses show approximate p-values based on simple, two-sided t-tests using bootstrapped standard errors. * p < 0.05, ** p < 0.01, *** p < 0.001

Table 5: Lasso-regularized coefficient estimates of the interaction model, 2009 German Parliamentary Election.

Note that by applying the Lasso method, pre-selection or other strategies for theory-testing, which are reserved for very well-specified causal mechanisms and are naturally prone to cover confounded effects, are not mandatory. In particular, also applied to a fully specified interaction model, our adjusted Lasso systematically selects the most important interaction terms. In order to demonstrate this and to identify all existing segment-specific reactions to issues, we additionally provide a model that includes all possible interactions between the issues and voter characteristics. This 'all-including-interactions model' and the corresponding unpenalized model can be found in Table 8 and 9 in Appendix B.

As can be immediately seen from Table 5, the shrinkage and regularization effect of the Lasso stabilizes this large model and efficiently and systematically selects important segment-specific reactions to issues. Therefore, the Lasso overcomes the two major drawbacks of interaction models – instability and exuberant complexity. In particular, half of

[42]The corresponding unpenalized model can be found in Table 7 in Appendix B.

the 32 specified interaction terms are efficiently removed from the model, and thus only 16 relevant interactions remain in the model.

Several interesting and large interaction terms have been selected by the Lasso. Take, for instance, the highly significant interaction effect between being Catholic and the issue of immigration on supporting the CDU. The issue of immigration has a significant positive impact on the CDU vote for voters having a Protestant religious affiliation, as can be seen from its main effect for this party. If we take instead the segment of Catholic voters, each unit-increase in the distance between them and the CDU on the issue of immigration increases the linear predictor of the CDU by only 0.02 ($0.36 - 0.38 = -0.02$; $-0.02 \cdot (-1) = 0.02$), suggesting that for Catholics attitudes on immigration do not strongly influence the vote decision in favor of the Christian Democrats. The significance of this interaction term indicates that there is a significant difference of the issue of immigration on the CDU vote between Catholics and Protestants, but it does not mean that the overall effect of the immigration issue on being Catholic is significant.

A series of further interesting interaction effects has been selected by the Lasso, including the interaction effect between the issue of immigration and political interest on the FDP vote. This negative effect indicates that the odds of FDP vote for the segment of politically less interested voters increases when voters disagree with the party stance on the issue of immigration.[43] Also note the selected interaction effect between the issue of nuclear energy and political interest on the CDU vote. The probability of voting for the CDU is more strongly influenced by the issue of nuclear energy among politically interested voters as among those being less politically interested. Notice that the interaction terms between the issue of taxes and the two considered measures of political sophistication are entirely removed from the model. The finding suggests that these voter segments do not differ in the importance they attach to the issue of taxes when they evaluate the parties on this issue. In addition, the main effects for political interest and having a high school degree are entirely removed from the party choice model. This important result shows that we observe only a difference between the segment of highly politically sophisticated and the less politically sophisticated ones when there is a disagreement between them and the respective party positions on the issues.

It can be concluded the Lasso considerably reduces the interacting structure and, thereby its interpretation. In particular, the findings demonstrate that, within the ML context, one would have to select important interactions manually or via elaborate testing procedures. By contrast, the Lasso approach easily allows including all possible interactions. As can be seen from Table 6, the Lasso-penalized interaction model also clearly outperforms the corresponding ML model. Using ML estimation, the inclusion of interactions leads to a remarkable deterioration of the model. By contrast, the Lasso-penalized interaction model exhibits by far the best crossvalidation score, AIC and BIC. Since the Lasso-regularized interaction model includes only 43 effects, compared to the 88 of the

[43] Coding of the covariate political interest: 1 (less interested), 0 (very interested).

corresponding ML model, it is much less complex and easier to interpret than the ML model.

Model	CV	AIC	BIC	$\widehat{\text{df}}$
ML interaction	184.88	1715.12	2129.11	88
Lasso interaction	172.81	1632.79	1819.98	43

Table 6: Comparison of interaction effects models based on predictive performance and complexity.

5 Conclusion

Multinomial logit models (MNLs) are a powerful tool to analyze electoral choice behavior in multiparty systems. Although offering the possibility to address new aspects of voting behavior, such as party-specific reactions to issue distances, the flexibility of these models implies at the same time an enormous proliferation of coefficients, highlighting the practical need for sophisticated parameter selection techniques.

In this paper, we address the problem of high-dimensional predictor spaces in models of party choice by introducing and outlining the benefits of regularization methods. We explicitly derive for the first time Lasso-type regularization techniques in the estimation of MNLs that are able to incorporate the alternative-wise specification of choice-specific covariates (e.g., issue distances). Since the Lasso is able to set parameters of irrelevant predictors to exactly zero, the corresponding effects are effectively removed from the model. Thus, the Lasso implicitly performs variable selection as a by-product of the estimation process, and therefore enforces parameter selection and reduction of the predictor space. Hence and in contrast to classical selection techniques, the Lasso guarantees continuous, stable and computationally efficient variable selection. We also illustrate that the usage of adaptive weights yields a considerable improvement of the Lasso's performance, both in terms of variable selection and prediction accuracy. Finally, we show that using a symmetric identifiability constraint prevents that the arbitrary choice of a reference category influences the selection of the most promising model. In particular, we demonstrate that the models' sparsity and complexity substantially vary across different choices of the reference category. Therefore, we strongly recommend the usage of symmetric side constraints as identifiability constraint in MNLs. Naturally, other interpretation techniques as well as the derivation of equilibria can build on these results. Even though the advantages of our Lasso approach are demonstrated by modeling electoral choices, it can be applied to any choice situation based on this type of model class facing the problem of parameter inflation.

By applying the Lasso method to the 2009 German Parliamentary Election, we demonstrate that our proposed approach massively reduces the model's complexity and simplifies its interpretation by selecting highly interesting and theoretically promising effects.

In addition, it improves the model's predictive performance. More specifically, Lasso-penalization clearly outperforms the simple ML estimator, for both the main effects and the interaction effects model.

A Properties of Lasso Estimates

Let $\hat{\boldsymbol{\theta}}_\lambda$ denote the Lasso estimator of the model's overall parameter vector $\boldsymbol{\theta}$ for tuning parameter λ. Let $\lambda_{\max} > \lambda_2 > \lambda_1 > 0$ and let $||\cdot||_1$ denote the L_1-norm.

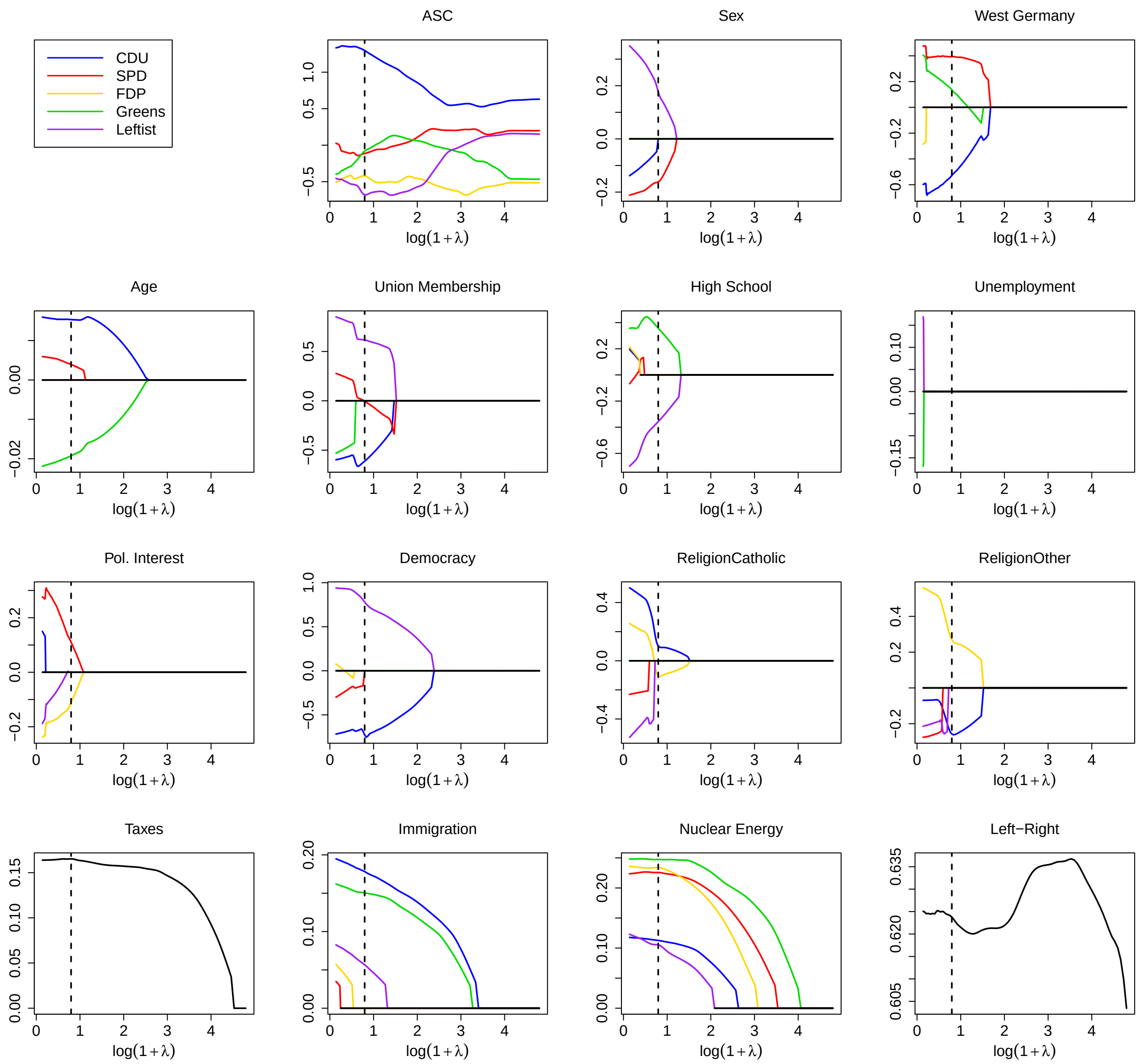
Then, it follows from the constraint definition of the Lasso that the L_1 norm of the Lasso is smaller than that of the ML estimator and that it decreases with increasing penalty level: $||\hat{\boldsymbol{\theta}}^{ML}||_1 = ||\hat{\boldsymbol{\theta}}_0||_1 > ||\hat{\boldsymbol{\theta}}_{\lambda_1}||_1 > ||\hat{\boldsymbol{\theta}}_{\lambda_2}||_1$. Note, however, that not every single coefficient is shrunk compared to its ML counterpart. For groups of highly correlated predictors, the Lasso typically only selects one of them while the others are set to zero. In such cases, the coefficient for the selected predictor partially subsumes the effects of the removed correlated predictors. Thus, in the presence of correlation among the predictors, single coefficients can get larger when λ is increased.

Due to its shrinkage property, the Lasso is a biased estimator. Its variance, by contrast, becomes smaller for a larger degree of penalization. It can be shown that $\text{Var}(\hat{\boldsymbol{\theta}}^{ML}) > \text{Var}(\hat{\boldsymbol{\theta}}_{\lambda_1}) > \text{Var}(\hat{\boldsymbol{\theta}}_{\lambda_2})$.[44]

The last property of the Lasso that we consider is its effective degrees of freedom. Since the Lasso is a shrinkage estimator, its effective degrees of freedom will intuitively be smaller than the number of estimated parameters. In Zou, Hastie, and Tibshirani (2007), it was shown that its effective degrees of freedom can be estimated by the number of nonzero parameters and that this df-estimator is unbiased and consistent: $\hat{\text{df}}(\hat{\boldsymbol{\theta}}_\lambda) = \sum_i I(|\hat{\theta}_{i,\lambda}| > 0)$. However, this general formula has to be handled with care due to the identifiability constraints used in MNLs. We give an explicit formula for the df of the Lasso for our model in Section 3.2.3.

[44]Here, $\text{Var}(\hat{\boldsymbol{\theta}})$ is a covariance matrix, so that the notation "$\text{Var}(\hat{\boldsymbol{\theta}}_{\lambda_1}) > \text{Var}(\hat{\boldsymbol{\theta}}_{\lambda_2})$" means that the difference of these matrices is positive definite.

B Extended Figures and Tables



Note: The coefficient paths are based on the main effects model in Table 2.

Figure 4: Coefficient paths for all predictors, 2009 German Parliamentary Election.

	CDU		SPD		FDP		Greens		Leftist	
ASC	1.87***	(0.000)	0.24	(0.582)	-0.58	(0.277)	-0.63	(0.308)	-0.91	(0.084)
Sex	-0.20	(0.255)	-0.26	(0.133)	0.02	(0.939)	0.09	(0.708)	0.35	(0.117)
West Germany	-0.61**	(0.005)	0.51*	(0.024)	-0.16	(0.565)	0.47	(0.122)	-0.22	(0.395)
Age	0.02***	(0.001)	0.01	(0.182)	-0.01	(0.675)	-0.02**	(0.004)	0.00	(0.996)
Union Membership	-0.50	(0.134)	0.36	(0.127)	-0.27	(0.421)	-0.56	(0.129)	0.96***	(0.001)
High School	0.22	(0.647)	-0.19	(0.673)	0.14	(0.785)	0.38	(0.454)	-0.55	(0.340)
Unemployment	-0.04	(0.937)	0.11	(0.830)	-0.31	(0.780)	-0.18	(0.795)	0.43	(0.394)
Pol. Interest	0.21	(0.587)	0.09	(0.820)	-1.22*	(0.015)	0.36	(0.495)	0.57	(0.200)
Democracy	-0.69***	(0.001)	-0.31	(0.092)	0.24	(0.301)	-0.22	(0.389)	0.98***	(0.001)
ReligionCatholic	-0.15	(0.709)	-0.41	(0.227)	1.21**	(0.007)	-0.11	(0.793)	-0.54	(0.178)
ReligionOther	-0.84*	(0.027)	-0.15	(0.678)	1.39**	(0.003)	0.15	(0.724)	-0.56	(0.191)
Taxes	0.19*	(0.018)	0.19*	(0.018)	0.19*	(0.018)	0.19*	(0.018)	0.19*	(0.018)
Immigration	0.40*	(0.015)	0.07	(0.588)	0.16	(0.297)	0.16	(0.368)	-0.06	(0.714)
Nuclear Energy	0.25*	(0.017)	0.35*	(0.049)	0.23	(0.196)	0.19	(0.167)	0.13	(0.382)
Left-Right	0.64***	(0.000)	0.64***	(0.000)	0.64***	(0.000)	0.64***	(0.000)	0.64***	(0.000)
Taxes X High School	0.06	(0.646)	0.06	(0.646)	0.06	(0.646)	0.06	(0.646)	0.06	(0.646)
Taxes X Pol. Interest	-0.06	(0.521)	-0.06	(0.521)	-0.06	(0.521)	-0.06	(0.521)	-0.06	(0.521)
Immigration X High School	-0.11	(0.465)	-0.05	(0.808)	-0.24	(0.238)	-0.15	(0.428)	0.09	(0.716)
Immigration X Pol. Interest	0.19	(0.136)	0.01	(0.963)	-0.37*	(0.029)	0.04	(0.789)	0.28	(0.057)
Nuclear Energy X High School	-0.04	(0.753)	-0.10	(0.610)	0.09	(0.727)	0.09	(0.566)	-0.12	(0.582)
Nuclear Energy X Pol. Interest	-0.17	(0.124)	-0.14	(0.446)	0.01	(0.970)	0.02	(0.874)	0.03	(0.865)
Immigration X ReligionCatholic	-0.43**	(0.010)	-0.12	(0.350)	0.32	(0.116)	0.00	(0.923)	-0.01	(0.947)
Immigration X ReligionOther	-0.36*	(0.028)	0.07	(0.629)	0.29	(0.100)	0.08	(0.625)	-0.10	(0.515)

Numbers in parentheses show approximate p-values based on simple, two-sided t-tests using bootstrapped standard errors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: ML coefficient estimates of the interaction model, 2009 German Parliamentary Election.

	CDU		SPD		FDP		Greens		Leftist	
ASC	1.51***	(0.000)	-0.13	(0.846)	-0.97	(0.124)	-0.14	(0.274)	-0.27	(0.459)
Sex	0		0		0		0		0	
West Germany	0		0		0		0		0	
Age	0		0		0		0		0	
Union Membership	-1.24*	(0.025)	0		0		0		1.24*	(0.017)
High School	0		0		0		0		0	
Unemployment	0		0		0		0		0	
Pol. Interest	0		0		0		0		0	
Democracy	-0.40	(0.083)	0		0.08	(0.596)	0		0.32	(0.221)
ReligionCatholic	-0.98*	(0.014)	0		1.20*	(0.010)	-0.22	(0.701)	0	
ReligionOther	-1.12**	(0.006)	0		1.04*	(0.032)	0.09	(0.444)	0	
Taxes	0.22	(0.052)	0.27	(0.060)	0		0.18	(0.369)	0.34	(0.087)
Immigration	0		0		0.12	(0.250)	0.32*	(0.048)	0	
Nuclear Energy	0.24	(0.083)	0.21	(0.245)	0.11	(0.309)	0.24	(0.176)	0.18	(0.265)
Left-Right	0.81***	(0.000)	0.56	(0.051)	0.44	(0.086)	0.28	(0.263)	0.60**	(0.004)
Taxes X Sex	0		-0.09	(0.238)	0.12	(0.254)	0		0	
Taxes X West Germany	0		0		0		0		0.19	(0.323)
Taxes X Age	0		0		0		0		0	
Taxes X Union Membership	0		0		0		-0.07	(0.328)	0.09	(0.343)
Taxes X High School	0		0		0		0		0	
Taxes X Unemployment	0		0		-0.59	(0.109)	0.20	(0.421)	0	
Taxes X Pol. Interest	0		0		0		0.20	(0.169)	0	
Taxes X Democracy	0		0		0.10	(0.294)	0		-0.26	(0.231)
Taxes X ReligionCatholic	-0.08	(0.557)	-0.08	(0.318)	0.33	(0.100)	-0.09	(0.424)	0	
Taxes X ReligionOther	-0.15	(0.186)	-0.19	(0.090)	0.07	(0.913)	-0.21	(0.165)	0	
Immigration X Sex	0		0		-0.07	(0.254)	0		0	
Immigration X West Germany	0.26	(0.106)	0		0.09	(0.387)	-0.16	(0.102)	-0.07	(0.218)
Immigration X Age	0		0		0		0.01*	(0.032)	0	
Immigration X Union Membership	0		0		-0.17	(0.234)	0.11	(0.204)	0	
Immigration X High School	0		0		-0.06	(0.279)	-0.05	(0.413)	0.14	(0.315)
Immigration X Unemployment	0		0.15	(0.377)	0		0		0.24	(0.272)
Immigration X Pol. Interest	0.24*	(0.046)	0		-0.17	(0.107)	0		0.14	(0.253)
Immigration X Democracy	0		0		0		0		0	
Immigration X ReligionCatholic	-0.34*	(0.045)	0		0		0		0.04	(0.468)
Immigration X ReligionOther	-0.18	(0.249)	0		0		0		-0.04	(0.597)
Nuclear Energy X Sex	0		0		0		0		-0.19	(0.085)
Nuclear Energy X West Germany	0		0		0		0		0	
Nuclear Energy X Age	0		0		0		0		0	
Nuclear Energy X Union Membership	0		-0.13	(0.260)	0.65	(0.156)	0		0	
Nuclear Energy X High School	0		0		0		0		0	
Nuclear Energy X Unemployment	0.11	(0.452)	0.50	(0.165)	1.52*	(0.017)	0		0	
Nuclear Energy X Pol. Interest	-0.13	(0.120)	0		0		0		0	
Nuclear Energy X Democracy	0.12	(0.243)	0		0		0		0.09	(0.306)
Nuclear Energy X ReligionCatholic	-0.11	(0.351)	0		0.15	(0.315)	0		0	
Nuclear Energy X ReligionOther	-0.13	(0.204)	0		-0.01	(0.668)	0		0	
Left-Right X Sex	0		0.31	(0.091)	0		-0.08	(0.270)	0	
Left-Right X West Germany	0		0		0		0		0	
Left-Right X Age	0		0		0		-0.01	(0.135)	0	
Left-Right X Union Membership	0		0		0		0.50	(0.098)	0.39	(0.093)
Left-Right X High School	0		0		0		0		0	
Left-Right X Unemployment	0		-1.04*	(0.037)	0.36	(0.113)	0		0	
Left-Right X Pol. Interest	0		0		0		0		0.13	(0.218)
Left-Right X Democracy	-0.07	(0.296)	0		-0.17	(0.322)	0		-0.16	(0.325)
Left-Right X ReligionCatholic	-0.29*	(0.045)	0.05	(0.757)	0.16	(0.584)	0.19	(0.562)	0.03	(0.875)
Left-Right X ReligionOther	-0.06	(0.586)	0.24	(0.229)	0.34	(0.088)	0.56*	(0.013)	-0.06	(0.527)

Numbers in parentheses show approximate p-values based on simple, two-sided t-tests using bootstrapped standard errors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Lasso coefficient estimates of the 'all-including-interactions model', 2009 German Parliamentary Election.

	CDU		SPD		FDP		Greens		Leftist	
ASC	2.90*	(0.024)	0.19	(0.871)	-2.01	(0.133)	-0.80	(0.558)	-0.28	(0.834)
Sex	-0.26	(0.671)	0.19	(0.739)	0.38	(0.588)	0.10	(0.887)	-0.42	(0.564)
West Germany	-0.61	(0.426)	1.11	(0.159)	0.47	(0.590)	-0.06	(0.946)	-0.91	(0.332)
Age	0.03	(0.093)	-0.01	(0.770)	-0.01	(0.612)	-0.01	(0.521)	0	
Union Membership	-1.62	(0.273)	-0.58	(0.599)	0.52	(0.817)	-0.89	(0.428)	2.57	(0.066)
High School	-0.07	(0.938)	-0.48	(0.602)	0.30	(0.755)	0.68	(0.464)	-0.44	(0.697)
Unemployment	2.29	(0.489)	-0.61	(0.769)	-4.51	(0.074)	0.36	(0.877)	2.47	(0.338)
Pol. Interest	0.43	(0.525)	-0.41	(0.569)	-1.73*	(0.031)	0.87	(0.329)	0.84	(0.316)
Democracy	-1.01	(0.149)	-0.25	(0.681)	0.87	(0.276)	-0.50	(0.504)	0.89	(0.345)
ReligionCatholic	-1.38	(0.105)	-0.57	(0.406)	2.26*	(0.030)	0.28	(0.735)	-0.58	(0.535)
ReligionOther	-1.81	(0.058)	-0.22	(0.786)	1.92*	(0.045)	0.91	(0.291)	-0.80	(0.422)
Taxes	0.71	(0.084)	0.57	(0.136)	-0.13	(0.745)	0.28	(0.588)	0.48	(0.308)
Immigration	0.14	(0.720)	0.08	(0.796)	0.17	(0.682)	0.40	(0.307)	0.35	(0.470)
Nuclear Energy	0.31	(0.327)	0.34	(0.425)	0.19	(0.646)	0.28	(0.462)	0.32	(0.455)
Left-Right	1.24*	(0.011)	0.61	(0.186)	0.52	(0.367)	0.29	(0.546)	0.58	(0.151)
Taxes X Sex	-0.19	(0.362)	-0.22	(0.338)	0.30	(0.219)	-0.18	(0.578)	0.10	(0.726)
Taxes X West Germany	-0.25	(0.413)	0.15	(0.579)	-0.03	(0.914)	0.13	(0.665)	0.23	(0.440)
Taxes X Age	0		0		-0.01	(0.453)	0.01	(0.364)	-0.01	(0.304)
Taxes X Union Membership	-0.28	(0.510)	-0.14	(0.690)	0.15	(0.861)	-0.46	(0.306)	0.34	(0.507)
Taxes X High School	0.07	(0.806)	-0.13	(0.738)	0.16	(0.664)	0.32	(0.314)	-0.06	(0.917)
Taxes X Unemployment	0.22	(0.856)	-0.39	(0.696)	-1.79	(0.086)	0.89	(0.416)	-0.31	(0.781)
Taxes X Pol. Interest	-0.09	(0.739)	-0.11	(0.671)	-0.10	(0.721)	0.37	(0.266)	-0.21	(0.440)
Taxes X Democracy	-0.01	(0.969)	-0.15	(0.514)	0.24	(0.368)	-0.08	(0.787)	-0.38	(0.335)
Taxes X ReligionCatholic	-0.18	(0.585)	-0.25	(0.291)	0.46	(0.187)	-0.26	(0.514)	0	
Taxes X ReligionOther	-0.42	(0.200)	-0.32	(0.270)	0.04	(0.902)	-0.32	(0.359)	0.08	(0.791)
Immigration X Sex	0.08	(0.733)	0.03	(0.857)	-0.16	(0.508)	0.20	(0.320)	-0.19	(0.365)
Immigration X West Germany	0.26	(0.302)	-0.09	(0.697)	0.24	(0.416)	-0.38	(0.158)	-0.33	(0.188)
Immigration X Age	0.01	(0.270)	0		0		0.01	(0.139)	0	
Immigration X Union Membership	-0.30	(0.589)	0.22	(0.488)	-0.41	(0.556)	0.32	(0.333)	-0.28	(0.527)
Immigration X High School	-0.05	(0.830)	0.01	(0.965)	-0.26	(0.399)	-0.15	(0.589)	0.14	(0.713)
Immigration X Unemployment	-0.21	(0.879)	0.26	(0.753)	-0.22	(0.813)	-0.40	(0.748)	0.84	(0.406)
Immigration X Pol. Interest	0.30	(0.171)	0.05	(0.813)	-0.37	(0.191)	0.11	(0.650)	0.22	(0.374)
Immigration X Democracy	-0.17	(0.455)	-0.10	(0.586)	0.12	(0.615)	-0.02	(0.907)	-0.13	(0.755)
Immigration X ReligionCatholic	-0.44	(0.116)	-0.07	(0.715)	0.04	(0.896)	0.01	(0.957)	0.11	(0.698)
Immigration X ReligionOther	-0.22	(0.433)	0.06	(0.820)	0.14	(0.616)	-0.14	(0.602)	-0.18	(0.565)
Nuclear Energy X Sex	0.08	(0.663)	0.03	(0.893)	-0.02	(0.932)	0.02	(0.942)	-0.48	(0.129)
Nuclear Energy X West Germany	0.08	(0.724)	0.13	(0.667)	-0.07	(0.763)	-0.05	(0.818)	0.05	(0.878)
Nuclear Energy X Age	0		0		0		0		0.01	(0.442)
Nuclear Energy X Union Membership	0.02	(0.963)	-0.29	(0.402)	1.21	(0.327)	-0.17	(0.627)	0.30	(0.573)
Nuclear Energy X High School	-0.02	(0.923)	-0.05	(0.870)	0.18	(0.628)	0.01	(0.971)	-0.03	(0.942)
Nuclear Energy X Unemployment	0.67	(0.437)	0.89	(0.449)	2.37	(0.067)	0.18	(0.907)	-0.19	(0.848)
Nuclear Energy X Pol. Interest	-0.23	(0.242)	-0.26	(0.404)	-0.03	(0.908)	-0.08	(0.732)	-0.06	(0.843)
Nuclear Energy X Democracy	0.20	(0.310)	0.15	(0.586)	0.08	(0.770)	-0.16	(0.487)	0.37	(0.241)
Nuclear Energy X ReligionCatholic	-0.22	(0.423)	-0.08	(0.745)	0.21	(0.525)	0.20	(0.481)	-0.36	(0.380)
Nuclear Energy X ReligionOther	-0.28	(0.285)	-0.12	(0.699)	-0.19	(0.444)	0.26	(0.304)	-0.12	(0.760)
Left-Right X Sex	-0.19	(0.396)	0.38	(0.212)	0.04	(0.920)	-0.36	(0.309)	0.09	(0.754)
Left-Right X West Germany	-0.10	(0.756)	0.25	(0.459)	0.21	(0.552)	0.25	(0.510)	-0.08	(0.812)
Left-Right X Age	0		-0.01	(0.411)	0		-0.01	(0.235)	0	
Left-Right X Union Membership	0.32	(0.535)	-0.19	(0.703)	0.45	(0.582)	0.62	(0.264)	0.77	(0.219)
Left-Right X High School	-0.12	(0.684)	-0.10	(0.819)	0.04	(0.927)	0.07	(0.864)	0.11	(0.822)
Left-Right X Unemployment	0.58	(0.614)	-1.54	(0.147)	1.31	(0.156)	0.25	(0.878)	0.58	(0.594)
Left-Right X Pol. Interest	0.11	(0.651)	-0.38	(0.301)	-0.30	(0.346)	-0.16	(0.650)	0.41	(0.178)
Left-Right X Democracy	-0.27	(0.297)	0.17	(0.581)	-0.27	(0.499)	0.11	(0.753)	-0.25	(0.503)
Left-Right X ReligionCatholic	-0.53	(0.119)	0.05	(0.874)	0.15	(0.768)	0.09	(0.809)	0.02	(0.961)
Left-Right X ReligionOther	-0.19	(0.586)	0.44	(0.238)	0.55	(0.130)	0.87*	(0.047)	-0.28	(0.409)

Numbers in parentheses show approximate p-values based on simple, two-sided t-tests using bootstrapped standard errors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 9: ML coefficient estimates of the 'all-including-interactions model', 2009 German Parliamentary Election.

References

Adams, James F., Samuel Merrill, and Bernard Grofman. 2005. *A unified theory of party competition: A cross-national analysis integrating spatial and behavioral factors.* Cambridge and New York: Cambridge University Press.

Alvarez, R. Michael, and Jonathan Nagler. 1998. "When politics and models collide: Estimating models of multiparty elections." *American Journal of Political Science* 42(1): 55–96.

Beck, Amir, and Marc Teboulle. 2009. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." *SIAM Journal on Imaging Sciences* 2(1): 183–202.

Ben-Akiva, Moshe E., and Steven R. Lerman. 1985. *Discrete choice analysis: Theory and application to travel demand.* Cambridge Massachusetts: MIT Press.

Boyd, Stephen P., and Lieven Vandenberghe. 2004. *Convex optimization.* Cambridge: Cambridge University Press.

Converse, Philip E. 1964. "The nature of belief systems in mass publics." In *Ideology and Discontent*, ed. David Apter. New York: Free Press , 240–268.

Dow, Jay K., and James W. Endersby. 2004. "Multinomial probit and multinomial logit: A comparison of choice models for voting research." *Electoral Studies* 23(1): 107–122.

Downs, Anthony. 1957. *An economic theory of democracy.* New York: Harper & Row.

Elff, Martin. 2007. "Social structure and electoral behavior in comparative perspective: The decline of social cleavages in Western Europe revisited." *Perspectives on Politics* 5(02): 277–294.

Elff, Martin. 2009. "Social divisions, party positions, and electoral behaviour." *Electoral Studies* 28(2): 297–308.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2010. "Regularization paths for generalized linear models via coordinate descent." *Journal of Statistical Software* 33(1): 1–22.

Gilljam, Mikael. 1997. "Symposium. The Directional Theory of Issue Voting: I The Directional Theory under the Magnifying Glass: A Reappraisal." *Journal of Theoretical Politics* 9(1): 5–12.

Green-Pedersen, Christoffer. 2007. "The Growing Importance of Issue Competition: The Changing Nature of Party Competition in Western Europe." *Political Studies* 55(3): 607–628.

Guinaudeau, Isabelle, and Simon Persico. 2014. "What is Issue Competition? Conflict, Consensus and Issue Ownership in Party Competition." *Journal of Elections, Public Opinion and Parties* 24(3): 312–333.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: Data mining, inference and prediction.* New York: Springer.

Hensher, David A., John M. Rose, and William H. Greene. 2007. *Applied Choice Analysis: A primer.* 1. publ., 3. print. ed. Cambridge: Cambridge University Press.

Hoerl, Arthur E., and Robert W. Kennard. 1970. "Ridge regression: Biased estimation for nonorthogonal problems." *Technometrics* 12(1): 55–67.

Krosnick, Jon A. 1990. "Government policy and citizen passion: A study of issue publics in contemporary America." *Political Behavior* 12(1): 59–92.

Lewis, Jeffrey B., and Gary King. 1999. "No evidence on directional vs. proximity voting." *Political analysis* 8(1): 21–33.

Louviere, Jordan J., David A. Hensher, and Joffre D. Swait. 2009. *Stated Choice Methods: Analysis and Applications.* 1. publ., 6. print. ed. Cambridge: Cambridge University Press.

Luskin, Robert C. 1987. "Measuring Political Sophistication." *American Journal of Political Science* 31(4): p 856–899.

Mauerer, Ingrid, Paul W. Thurner, and Marc Debus. 2015. "Under Which Conditions Do Parties Attract Voters' Reactions to Issues? Party-Varying Issue Voting in German Elections 1987–2009." *West European Politics* , 1–23.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behaviour." In *Frontiers in Econometrics*, ed. Paul Zarembka. New York: Academic Press , 105–42.

Mebane, Walter R., John E. Jackson, and Jonathan Wall. 2014. "Preference heterogeneities in models of electoral behavior." *Working paper presented at the 2014 Annual Meeting of the Midwest Political Science Association, Chicago, April 3-6, 2014* .

Meguid, Bonnie M. 2005. "Competition between Unequals: The Role of Mainstream Party Strategy in Niche Party Success." *The American Political Science Review* 99(3): 347–359.

Merrill III, Samuel, and Bernard Grofman. 1997. "Response to Macdonald and Rabinowitz." *Journal of Theoretical Politics* 9(1): 57–60.

Rovny, Jan. 2012. "Who Emphasizes and who Blurs? Party Strategies in Multidimensional Competition." *European Union Politics* 13(2): 269–292.

Schofield, Normal, Andrew D. Martin, Kevin M. Quinn, and Andrew B. Whitford. 1998. "Multiparty electoral competition in the Netherlands and Germany: A model based on multinomial probit." In *Empirical Studies in Comparative Politics*, eds. Melvin J. Hinich, and Michael C. Munger. Springer US , 39–75.

Singh, Shane. 2014. "Linear and quadratic utility loss functions in voting behavior research." *Journal of Theoretical Politics* 26(1): 35–58.

Thurner, Paul W. 2000. "The empirical application of the spatial theory of voting in multiparty systems with random utility models." *Electoral Studies* 19(4): 493–517.

Tibshirani, Robert. 1996. "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society B* 58(1): 267–288.

Train, Kenneth. 2009. *Discrete choice methods with simulation.* 2 ed. Cambridge University Press.

Tutz, G., W. Pößnecker, and L. Uhlmann. 2015. "Variable Selection in General Multinomial Logit Models." *Computational Statistics & Data Analysis* 82: 207–222.

Tutz, Gerhard. 2010. "Guest Editorial: Regularisation methods in regression and classification." *Statistics and Computing* 20(2): 117–118.

Tutz, Gerhard. 2012. *Regression for categorical data.* Cambridge University Press.

Tutz, Gerhard, and Gunther Schauberger. 2013. "Visualization of categorical response models - from Data Glyphs to Parameter Glyphs." *Journal of Computational and Graphical Statistics* 22: 156–177.

Wagner, Markus. 2012. "When do Parties Emphasise Extreme Positions? How Strategic Incentives for Policy Differentiation Influence Issue Importance." *European Journal of Political Research* 51(1): 64–88.

Westholm, Anders. 1997. "Distance versus direction: The illusory defeat of the proximity theory of electoral choice." *American Political Science Review* 91(04): 865–883.

Zou, Hui. 2006. "The adaptive lasso and its oracle properties." *Journal of the American Statistical Association* 101(476): 1418–1429.

Zou, Hui, Trevor Hastie, and Robert Tibshirani. 2007. "On the "degrees of freedom" of the lasso." *The Annals of Statistics* 35(5): 2173–2192.