

BLP-Lasso for Aggregate Discrete Choice Models Applied to Elections with Rich Demographic Covariates*

Benjamin J. Gillen

California Institute of Technology

Hyungsik Roger Moon

University of Southern California

Sergio Montero

California Institute of Technology

Matthew Shum

California Institute of Technology

October 15, 2015

Abstract

Economists often study consumers' aggregate behavior across markets choosing from a menu of differentiated products. In this analysis, local demographic characteristics can serve as controls for market-specific heterogeneity in product preferences. Given rich demographic data, implementing these models requires specifying which variables to include in the analysis, an ad hoc process typically guided primarily by a researcher's intuition. We propose a data-driven approach to estimate these models applying penalized estimation algorithms imported from the machine learning literature along with confidence intervals that are robust to variable selection. Our application explores the effect of campaign spending on vote shares in data from Mexican elections.

Keywords. Demand estimation, Elections, Post-model selection inference, Lasso, Big data

*We owe special thanks to Alexander Charles Smith for important insights early in developing the project. Staff at INE and INEGI were remarkably helpful in obtaining the data. We are grateful for comments from David Brownstone, Martin Burda, Garland Durham, Jeremy Fox, Gautam Gowriskaran, Chris Hansen, Stefan Holderlein, Ivan Jeliazkov, Dale Poirier, Guillame Weisang, Frank Windmeijer, and seminar participants at the Advances in Econometrics Conference on Bayesian Model Comparison, the California Econometrics Conference, Stanford SITE conference on Empirical Implementation of Theoretical Models of Strategic Interaction and Dynamic Behavior, the ASSA's Annual Meetings, UC Irvine, Cal Poly San Luis Obispo, and the University of Arizona.

Contents

1	Introduction	1
2	Related Literature	3
2.1	Model Selection and Inference	4
2.2	Structural Models of Campaign Spending and Voting	6
3	Voting in Mexico: Parties and Preferences	8
3.1	Fundraising and Advertising in Mexican Elections	9
3.2	Parties & Coalitions in the 2012 Chamber of Deputies Election	10
3.3	Electoral and Census Data	11
4	Model LM-F: Homogeneous Voters with Fixed Controls	14
4.1	The Structural Model and Estimation Strategy	15
4.2	A Preselected Model for Returns to Campaign Spending	18
5	Model LM-S: Variable Selection and Inference	20
5.1	Sparsity Assumptions and Regularity Conditions	20
5.2	Selecting Control Variables for Inference	22
5.3	Post-Selection Estimation and Inference	24
5.4	Returns to Campaign Expenditures after Control Selection	25
6	Model RC-F: Heterogeneous Impressionability with Fixed Controls	27
6.1	GMM Estimation with the BLP Model	28
6.2	Campaign Spending with Heterogeneous Impressionability	30
7	Model RC-S: Variable Selection in the BLP Voting Model	31
7.1	Sparsity Assumptions and Regularity Conditions	32
7.2	Implementing Variable Selection via Penalized GMM	34
7.3	Post-Selection Inference via Unpenalized GMM	37
7.4	Heterogeneous Impressionability after Control Selection	39
8	Generalizations and Extensions	41
9	Conclusion	42
	References	43
	Appendices	48
A	Iterative Computation for Penalty Loadings	48
A.1	Iterative Computation for Linear Models	48
A.2	Iterative Computation for Nonlinear Models	49

A.3	GMM Penalty for Verifying First Order Conditions	51
B	Detailed Statements of Model Assumptions	52
B.1	Notation	52
B.2	Assumption 2: Exact Sparsity in Preferences and Spending	52
B.3	Assumption 3: Regularity Conditions for High-Dimensional Linear Logit . .	53
B.4	Assumption 5: Regularity Conditions for GMM Estimator	54
B.5	Assumption 6: Sparsity Assumptions for High-Dimensional BLP Model . . .	54

1 Introduction

When analyzing aggregated data about consumers' choices in different regional markets, researchers must account for the demographic characteristics of local markets that might drive observable variability in consumers' preferences and firms' pricing policies. The abundance of such variables, whether from census data, localized search trends, or local media viewership surveys, immediately confronts researchers defining which model to adopt for this analysis with difficult questions. Which variables should be included in the model? Which controls can be excluded from the analysis without introducing omitted variable bias? How sensitive are the estimated effects of a firm's pricing policy on their market share to these specification decisions?

In the current paper, we hope to help answer these questions by providing data-driven algorithms for addressing model selection in analyzing consumer demand data. Our main contribution here is to apply recent econometric results from the variable selection literature to a popular nonlinear aggregate demand model. Specifically, our technique generalizes procedures from Belloni et al. (2012) and Belloni et al. (2013a) for selecting variables to a nonlinear Berry et al. (1995) model of consumer demand with random coefficients. Our asymptotic results extend Gillen et al. (2014)'s results by adopting techniques from Fan and Liao (2014)'s analysis of penalized GMM estimators, particularly the conditions for an oracle property that ensures all necessary variables are included in the model.

The specific problem of interest addresses high-dimensional demographic data for local markets that may help characterize local preferences. To address this problem, we adopt techniques proposed by the literature on machine learning to identify the demographic characteristics that exert the most important influence on observed market shares. As we discuss in section 2.1, these innovative algorithms present powerful devices for variable selection that require some care in their implementation. When properly deployed through multiple iterations of variable selection with appropriate penalization, these algorithms identify all the variables necessary for valid inference in the model.

We conduct an empirical investigation of campaign expenditures' influence on election

outcomes, utilizing structural models inspired by the industrial organization and marketing literatures. In elections, the “consumers” are the voters, the “market” is the voting district, and the set of available “products” is the set of political parties running in the district. We use our technique to analyze the impact of campaign expenditures on candidate vote shares in Mexican elections. With access to the full census records for Mexico, we have rich demographic data for each voting district. We also have some variability in the “market structure” or the set of political parties competing in each district, since Mexican elections allow parties to form partial coalitions. In districts where parties coordinate, the number of competing candidates is smaller than in those districts where they compete. Our analysis yields the robust finding that campaign expenditures significantly influence voter preferences.

We present our inferential technique in a series of four progressively more complex models. We introduce the simple discrete choice approach to testing the influence of campaign expenditures on voting in section 4. This simple setting assumes voters vote for their most preferred candidate, with voter preferences represented by a linear utility function. This simplicity admits a standard generalized linear model, which we refer to as Model LM-F (Linear Model with Fixed Controls), for candidate vote shares, providing a first look at the influence of campaign expenditures on vote shares in Mexican elections with a pre-selected model. The results show that campaign expenditures make a positive and significant contribution to a candidate’s vote share, though this contribution is mitigated by diminishing marginal returns.

We then introduce data-driven variable selection for the demographic controls we include in the model. In so doing, section 5 applies the techniques proposed in Belloni et al. (2012) and Belloni et al. (2013b) to develop our second empirical specification, Model LM-S (Linear Model with Selected Controls). In this discussion, we explicitly present the sparsity assumptions required for consistent and valid inference and describe the exact algorithm for performing that analysis. Implementing Model LM-S for the Mexican voting data, we illustrate how the algorithm can provide an agnostic characterization of the robustness of our results to model specification. In particular, the variable selection algorithm is governed by

only two “tuning” parameters, both of which are constrained by theory to lie in a reasonably small band of candidate values. Varying these tuning parameters allows us to verify the robustness of our empirical findings from Model LM-F to different specifications.

Next, Model RC-F (Random Coefficients with Fixed Controls) allows heterogeneity in voters’ responses to campaign expenditures in section 6. This specification corresponds to the “BLP” random coefficients logit model (Berry et al. (1995)), which is a workhorse model in empirical industrial organization. We fit Model RC-F to Mexican data using the same pre-specified set of demographic variables for controls used in the linear Model LM-F. Though we find very little evidence of heterogeneous impressionability, the impact of campaign expenditures on vote shares from the linear specification remains robust.

Our main innovation lies in the final specification, Model RC-S (Random Coefficients with Selected Controls). This model extends the results from Gillen et al. (2014) to a setting where the number of potential parameters grows exponentially with the sample size. Estimation and inference here poses some conceptual and computational challenges. The analytical development incorporates asymptotic results from Fan and Liao (2014), which builds on Caner and Zhang (2013) and Belloni et al. (2012) by establishing an oracle property for penalized GMM estimators in the ultra-high dimensional setting. Computationally, we start by fitting Model RC-F with the variables selected by Model LM-S to recover latent mean utilities and optimal instruments from the nonlinear BLP voting model. We then select a series of additional variables for robust inference, specifically including controls for observable heterogeneity in the optimal instruments. Finally, we verify the first order conditions of the penalized GMM-estimator to ensure we find a local optimum to which we can apply Fan and Liao (2014)’s oracle property.

2 Related Literature

The current paper sits at the intersection of political science, economics, and statistics. Our application addresses a well-worn question on how expenditures by a political campaign

influence the outcome of an election. The inferential model we use to investigate this question is grounded in structural econometric methods used for consumer demand estimation by researchers in industrial organization and marketing. Finally, the statistical techniques we apply utilize recent innovations in machine learning developing automated techniques for variable selection.

2.1 Model Selection and Inference

Data-driven approaches to variable selection represents one of the most active areas of statistical research today. Tibshirani (1996)'s Lasso estimator ushered in a new approach to estimation in high-dimensional settings by incorporating convex penalties to least-squares objective functions. The penalized estimation technique has been further developed by Fan and Li (2001)'s SCAD penalty, Zou and Hastie (2005)'s elastic net and Huang et al. (2008)'s Bridge estimator, Bickel et al. (2009)'s infeasible lasso, and Zhang (2010)'s minimax concave penalty. This literature has also inspired several closely related estimators, including Candes and Tao (2007)'s Dantzig selector and Gautier and Tsybakov (2011)'s feasible Dantzig selector as well as Belloni et al. (2011)'s Square-Root Lasso. Each of these estimators incorporate some form of $L1$ -regularization to the objective function's maximization problem, selecting variables for the model by imposing a large number of zero coefficients on the solution.

For an estimator that imposes a large number of zero coefficients in the solution to be consistent, it must be the case that a large number of zero coefficients are present in true model for the data generating process. This restriction on the true parameters of the model takes the form of a sparsity assumption. In its early formulations, the sparsity restriction was stated as an upper bound on the $L0$ or $L1$ norm of the true coefficients.¹ If an estimator classifies zero and non-zero coefficients with perfect accuracy as the sample grows, the estimator satisfies an oracle property. In order to establish an oracle property, the sparsity

¹Generalized notions of sparsity appear in Zhang and Huang (2008) and Horowitz and Huang (2010), which allow for local perturbations in which the zero-coefficients are very small. A similar approach appears in Belloni et al. (2012) and Belloni et al. (2013a) characterizing inference under an approximate sparsity condition that constrains the error in a sparse representation of the true data generating process.

restrictions need to be coupled with a minimum absolute value for non-zero coefficients to ensure they are selected by the penalized estimator. Intuitively, the variability of a single residual (which could be explained by an erroneously included explanatory variable) needs to be dominated by the penalty, which in turn needs to be dominated by the effect of a non-zero regressor (to justify the penalty associated with the coefficient’s non-zero value).

Performing inference after model selection, even with an estimator that satisfies the oracle property, has presented a non-trivial challenge to interpreting the results of estimators that incorporate these techniques. Leeb and Pötscher (2005, 2006, 2008) present early critiques of the sampling properties for naïvely-constructed test statistics after model selection, illustrating the failure of asymptotic normality to hold uniformly and the fragility of the bootstrap for computing standard errors in the selected model. Lockhart et al. (2014), published with a series of comments, propose significance tests for lasso estimators that perform well on “large” coefficients but are less effective for potentially “small” coefficients for which the significance tests are not pivotal due to the randomness of the null hypothesis. In a series of papers, Belloni et al. (2013a) and Belloni et al. (2012) propose techniques for inference on treatment effects in linear, instrumental variables, and logistic regression problems. These techniques incorporate multiple stages of variable selection with data-driven penalties that ensure the relevant controls are included in the econometric model before performing inference in an unpenalized post-selection model. By focusing on inference for a predefined, fixed-dimensional, subset of coefficients, the selected models represent a desparsified data generating process, with inference results from Van de Geer et al. (2014) providing uniformly valid confidence intervals.

Extending these techniques from least squares regression models to more general settings presents additional challenges. Fan and Li (2001), Zou and Li (2008), Bradic et al. (2011), and Fan and Lv (2011) propose methods for analyzing models defined by quasi-likelihood. Our application focuses on GMM estimators, whose properties in high-dimensions are considered by Caner (2009), Caner and Zhang (2013), Liao (2013), Cheng and Liao (2015), and Fan and Liao (2014). Several of these papers address the issue of moment selection, as in

Andrews (1999) and Andrews and Lu (2001). As our application considers an environment with a fixed set of instruments, our analysis does not require moment selection but makes heavy use of the oracle properties established by Fan and Liao (2014).

Our model builds directly on Gillen et al. (2014)’s analysis of demand models with complex products. The Gillen et al. (2014) application considers aggregate demand models where the dimension of the vector of product characteristics is large, on the same order of magnitude as the number of observations. Our current application utilizes variable selection to mitigate an incidental parameter problem in characterizing voter preferences. This utilization is similar in motivation to Harding and Lamarche (2015), who use a penalized quantile regression to allow for heterogeneity in individual nutritional preferences when analyzing a household grocery consumption data. In addition to applying Gillen et al. (2014)’s approach to an interactive fixed-effects model, our analysis allows for material weaker conditions on the data generating process than in Gillen et al. (2014). Notably, by applying Fan and Liao (2014)’s oracle properties, we can allow for the number of parameters to grow exponentially with the number of observations, rather than linearly.

2.2 Structural Models of Campaign Spending and Voting

Empirical analysis of voting data presents a particularly challenging exercise for political scientists due to the large number of factors driving voter behavior, endogeneity induced by party competition and candidate selection, and behavioral phenomena driving individual voter decisions. Including early work from Rothschild (1978) and Jacobson (1978), a number of political scientists have explored the effect of campaign spending on aggregate vote shares, often coming to different conclusions on its importance in influencing vote share by informing, motivating, and persuading voters. These inconclusive results arise in part due to challenges in identifying valid and relevant instruments (Jacobson, 1985; Green and Krasno, 1988; Gerber, 1998). Gordon et al. (2012) discuss several challenges to this research agenda, highlighting the value of incorporating historically underutilized empirical methods from marketing researchers.

A nascent literature in political science adopts structural approaches to inference for analyzing political data. Discrete choice approaches to analyzing voting data date back to Poole and Rosenthal (1985) and King (1997). Among the early adopters of this approach are Che et al. (2007), who utilize a nested logit model that takes advantage of individual voter data to identify the impact of advertisement exposure on their behavior. The problem we consider is closest to Rekkas (2007), Milligan and Rekkas (2008), and Gordon and Hartmann (2013), who apply a Berry et al. (1995) model to infer the impact of campaign expenditures on aggregate voting data. The analysis presented in Gordon and Hartmann (2013) provides an excellent motivation for our proposed inference technique. Though they find a robust evidence that campaign spending on advertisement positively contributes to a candidate’s vote share, the magnitude of this contribution varies by a factor of 3 depending on the specification of controls adopted. Their extremely large sample allows them to adopt very rich models of fixed effects and, in the most flexible models, the significance of the contribution of campaign expenditures to vote shares drops to 10%. A natural concern is that this loss of significance is in part due to an excessively conservative model of control variables. Our data-driven approach to selecting these control variables provides an agnostic approach to addressing some of the inherent ambiguity in determining which of these estimates is “most correct.”

A number of other researchers have also adopted a structural approach to analyze voting behavior. Degan and Merlo (2011) present a structural model for analyzing multiple concurrent elections in US Congressional and Presidential elections, with an extension by Levonyan (2013) analyzing the influence of the Presidential election on the outcome of California’s Proposition 8. Kawai and Watanabe (2013) adopt a structural approach in investigating strategic voting behavior using Japanese general election data. Kawai (2014) provides a dynamic extension of Erikson and Palfrey (2000)’s model of fund-raising and campaigning to analyze elections for US House of Representatives while adopting a control function approach to mitigate unobserved heterogeneity in voter behavior. Our application is clearly most closely related to Montero (2015)’s structural analysis of the incentives for coalition

formation in Mexican elections.

Beyond structural approaches for analyzing equilibrium outcomes, a massive body of empirical research investigates the influence of campaign expenditures on vote shares using natural and field experiments. These investigations are particularly valuable in their ability to differentiate how different styles of campaign advertising influences voter behavior. Gerber (2011) surveys much of this literature. Though our inference technique is derived in the context of a structural model of voting, the approach to selecting demographic control variables could be readily adopted to these environments.

3 Voting in Mexico: Parties and Preferences

Mexico is a federal republic with the executive branch headed by the president and legislative power wielded by a bicameral Congress. Our focus is restricted to elections for the lower chamber of Congress, known as the Chamber of Deputies, which are held every three years. Mexico is split into 300 electoral districts (see Figure 1), with the current boundaries drawn by the federal electoral authority, the Federal Electoral Institute (*Instituto Federal Electoral*, IFE), in 2005.² These boundaries were set with the objective of equalizing population subject to additional constraints preserving state boundaries and representation. While voting is mandatory in these elections for all citizens aged 18 and older, there are no sanctions enforcing participation. Legislators can be re-elected only in non-consecutive terms, limiting the incumbency power in these elections.

The Chamber of Deputies has 500 members, 300 of whom directly represent a district after being elected by a simple plurality of votes in a direct ballot. The candidates in these district races can be nominated by a single party or as a representative of a coalition of multiple political parties. Election laws allow candidates to run independently or as a write-in campaign, but their vote shares are negligible. The remaining 200 seats in the chamber are allocated according to a proportional representation (PR) rule. Specifically, the votes are

²In March 2014, IFE was transformed into the National Electoral Institute (*Instituto Nacional Electoral*, INE).

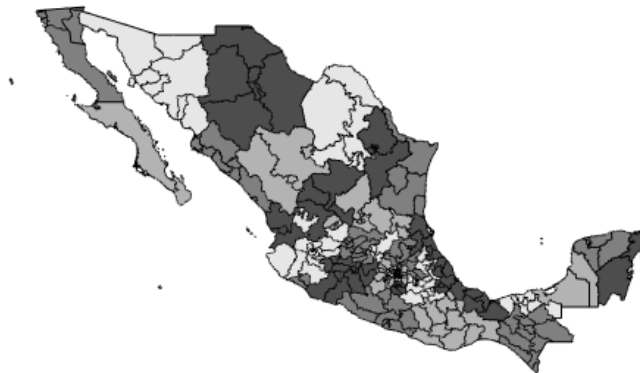


Figure 1: Mexican states (shaded) and electoral districts (delimited)

pooled by party across all districts and each party receives a share of the 200 proportional to the number of votes received by that party's candidates.³ To identify candidates for the PR assignment, parties submit national lists of up to 200 candidates concurrent with registering district candidates. Parties must secure at least 2% of the national vote for accreditation to hold seats in the legislature, with the votes for unaccredited parties annulled.

3.1 Fundraising and Advertising in Mexican Elections

Campaign funding for Mexican parties is allocated from the federal budget⁴ Of this total allocation, 30% is divided equally among all the parties, with the remaining 70% distributed in proportion to the parties' national vote shares in the most recent Chamber of Deputies election. The electoral authority then caps funding from other sources to 2% of the year's total public funding, ensuring public funds serve as the primary source of party incomes. Consequently, candidate fundraising is not prominent in these elections, with the party national committees mainly supplying the financial and administrative resources to run individual

³Additional restrictions for the PR assignment preclude any party from getting more than 300 total seats in the chamber or a share of seats that exceeds the party's national vote share by over 8 percentage points. In these cases, the excess PR seats are divided among the remaining parties proportionally to vote shares, though the adjustment is carried out only once: after performing the adjustment, if a new party exceeds the 8-percentage-points restriction with its additional share of seats, the process does not iterate.

⁴The total equals 65% of Mexico City's legal daily minimum wage multiplied by the total number of registered voters. After converting from Mexican *pesos* to U.S. dollars, this funding totaled about US\$250 million in 2012.

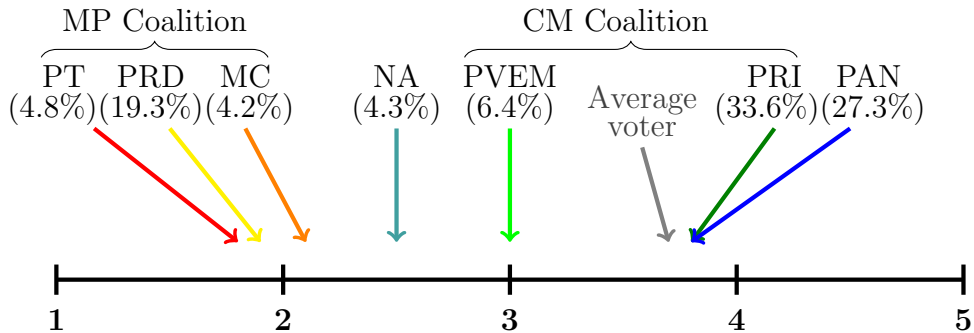
campaigns.

Campaigns take place within a fixed window of time: they must end 3 days before the day of the election and can only last up to 90 days in Presidential election years and 60 days in intermediate election years. Campaign advertising media is highly restricted in Mexican elections. The only legal access to TV and radio advertising is provided by the electoral authority to the parties free of charge. The total airtime is fixed and distributed to parties similarly to the public funding, with 30% divided equally and the remaining 70% proportionally to the parties' national vote shares in the most Chamber of Deputies election.

3.2 Parties & Coalitions in the 2012 Chamber of Deputies Election

The 2012 Chamber of Deputies election included seven political parties, two of which participated independently, three of which formed a total coalition, and two of which formed a partial coalition. The National Action Party (*Partido Acción Nacional*, PAN) and the New Alliance Party (*Partido Nueva Alianza*, NA) participated independently. The total coalition was called the Progressive Movement (*Movimiento Progresista*, MP), including the Party of the Democratic Revolution (*Partido de la Revolución Democrática*, PRD), the Labor Party (*Partido del Trabajo*, PT), and the Citizens' Movement (*Movimiento Ciudadano*, MC). The partial coalition was called the Commitment for Mexico (*Compromiso por México*, CM) and consisted of the Institutional Revolutionary Party (*Partido Revolucionario Institucional*, PRI) and the Ecologist Green Party of Mexico (*Partido Verde Ecologista de México*, PVEM). The partial coalition coordinated on a single candidate in 199 of the 300 districts races, 156 of which featured a jointly nominated PRI candidate, and the remaining 43, a joint PVEM candidate.

Figure 2 presents the parties on a one-dimensional ideology spectrum based on a national poll by Consulta Mitofsky in 2012, along with their national vote shares in the 2012 election. In analyzing the data on vote shares, we will treat coalitions as a single party. As such, elections consist of five parties in those districts where PRI and PVEM candidates ran separately and four parties where one candidate runs as part of the CM coalition. We do



Source: Consulta Mitofsky (2012). One thousand registered voters were asked in December of 2012 to place the parties and themselves on a five-point, left-right ideology scale. Arrows point to national averages. Parties' vote shares in parentheses.

Figure 2: Left-right ideological identification of Mexican parties and voters

include the actual party affiliation of the coalition candidate as a characteristic.

3.3 Electoral and Census Data

We utilize the data analyzed by Montero (2015)'s investigation of the coalition formation incentives of political parties. Individual-level voting data are not available as votes are cast anonymously, but district-level vote totals are publicly available online from the electoral authority. The final composition of the Chamber of Deputies, including the PR seats, is presented in Table 1.

To control for observable heterogeneity in voter preferences by district, we have access to rich demographic data—over 200 variables—from the 2010 population census, which the National Statistics and Geography Institute (*Instituto Nacional de Estadística y Geografía*, INEGI) makes available on a geo-electoral scale. Table 2 highlights summary statistics for districts' demographic composition along four specific variables regarding female heads of households, population over 64, education, and economic well-being. Breaking these characteristics down by the coalition structure of the parties within the district doesn't clearly indicate that demographics drive the coalition decision.

Campaign spending data are self-reported by the parties to the electoral authority, subject

Table 1: Chamber of Deputies composition after 2012 election

Party	Direct representation seats	Proportional representation seats	Total
PRI	158	49	207
PVEM	19	15	34
PAN	52	62	114
MP	71	64	135
NA		10	10
Total	300	200	500

Table 2: District characteristics

Variable	Districts with distinct PRI, PVEM candidates		Districts with joint CM candidate	
	Mean	Std. dev.	Mean	Std. dev.
Female head of household (% of total)	23.8	3.1	25.1	4.6
Pop. over 64 (% of over 17)	10.6	2.6	9.6	2.7
Avg. years of schooling (among over 14)	7.8	1.3	8.5	1.5
Household owns a car (% of total)	45.3	17.7	42.5	14.8

to audits. Audited spending data for the 2012 election are not yet available. We ignore misreporting as a source of measurement error, but, for comparison, campaign spending was *overreported* by about 4% in 2006, while no discrepancies were found in 2003. We focus on total spending per candidate, since we do not have access to detailed information on how funds were allocated to different forms of campaigning.

Table 3 reports average spending in a district by parties, broken down by the coalition structure of that district. To characterize the geographic dispersion of campaign spending, Figure 3 maps each party’s expenditures. We note that there is substantial variation in campaign spending by parties across neighboring districts, indicating that parties’ spending decisions are made strategically for each district. Specifically, the variability in expenditures between parties within a district is greater than would be expected by differences in the price of media, which would affect all parties equally.

Table 3: Campaign spending (in thousands of U.S. dollars)

Party	Districts with distinct PRI, PVEM candidates		Districts with joint CM candidate	
	Mean	Std. dev.	Mean	Std. dev.
PRI	54.9	11.0		
PVEM	18.3	7.6		
CM			83.5	31.2
PAN	38.0	10.4	42.1	13.1
MP	56.4	19.7	55.4	12.3
NA	19.7	8.5	17.2	5.6

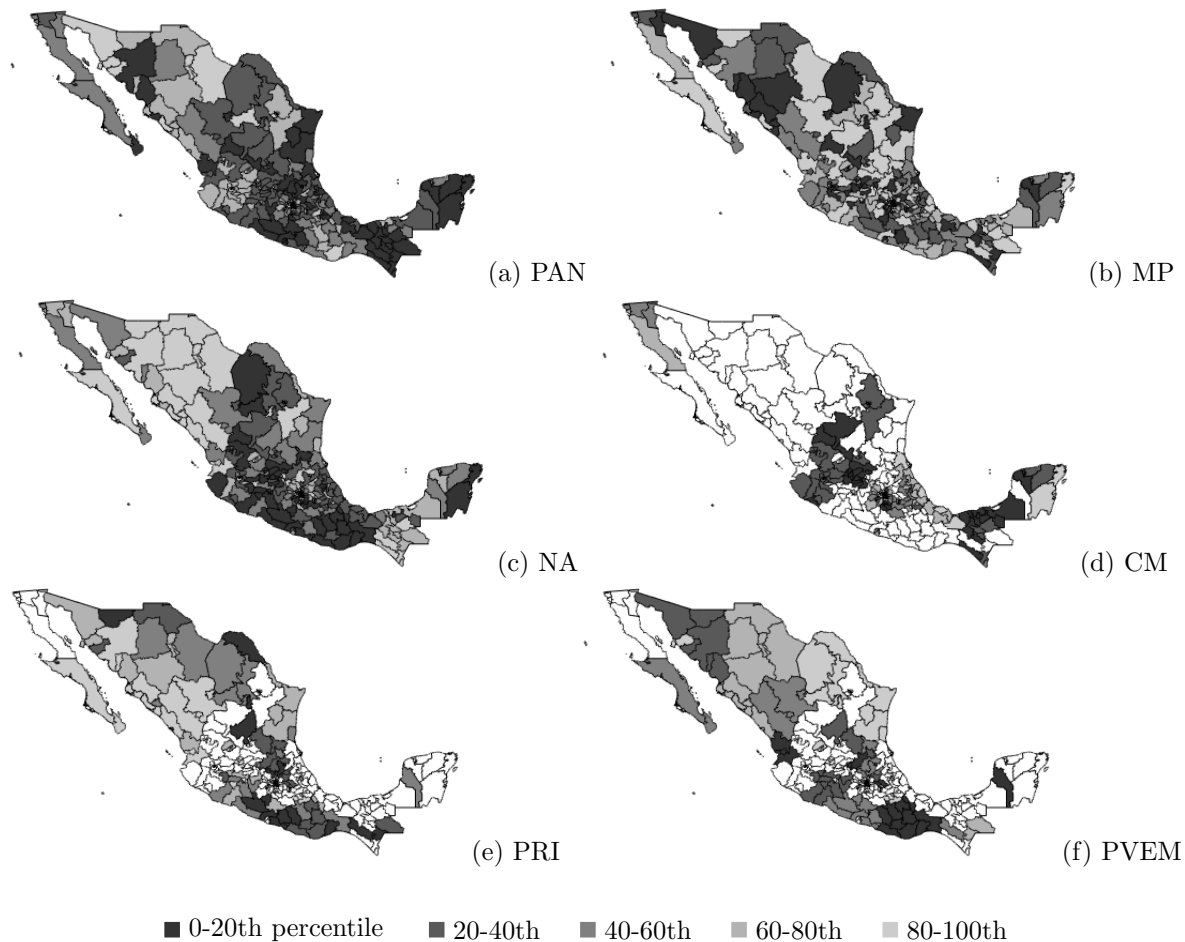


Figure 3: Geographic distribution of campaign spending by party

4 Model LM-F: Homogeneous Voters with Fixed Controls

We begin by introducing the structural model for voting in a setting free of complication by variable selection and nonlinear effects. This approach allows us to describe the economic environment for voting decisions and preferences before focusing on the methodological issues introduced by model specification tests. We then estimate a pre-selected model for control variables and instruments for the voting data in Mexico.

4.1 The Structural Model and Estimation Strategy

In district t , we observe vote shares based on individual voters (indexed by i) who choose from among the candidates competing in the district (indexed by $j = 1, \dots, J$). We represent the option to not vote or to write in a non-party candidate as an “outside good” indexed by $j = 0$. To characterize preferences for a representative voter in the district, we observe a vector of K_0 demographic characteristics for the district, denoted x_{0t} and K_1 characteristics describing the candidate for party j in that district, denoted x_{1jt} . The endogenous treatment variable of interest, campaign spending in the district by a candidate, is represented by p_{jt} .⁵ Finally, we allow exogenous unmodelled variation in voters’ preferences through a product-market specific latent shock, ξ_{jt} .

We begin by introducing preferences to our model in a restrictively homogeneous setting without random effects in preference characteristics, though we will relax these assumptions later. In district t , suppose consumer preferences are homogeneous up to an idiosyncratic, individual specific shock, denoted ϵ_{ijt} . This simplification allows us to represent consumer i ’s latent utility from voting for candidate j :

$$u_{ijt} = x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt} + \epsilon_{ijt}. \quad (1)$$

Note that, though the candidate-specific characteristics influence voter preferences in a common way across parties, district-specific demographics act as interactive fixed-effects, impacting voter preferences differently for different parties. The latter form of heterogeneity allows national party platform positions to influence local voting preferences depending on the district’s demographic composition.

Assuming the individual ϵ_{ijt} shocks are independently distributed with a Type-I Extreme Value distribution and normalizing the utility of not voting to 0, the probability of a

⁵For expositional purposes, we treat p_{jt} as a scalar, though it could be interpreted as a fixed-dimensional vector of treatment variables. Our empirical specification will allow for campaign expenditures to exert both a linear and quadratic influence on voter latent utilities.

randomly-selected district t voter choosing candidate j is given by the usual logit form:

$$Pr \{y_{ijt} = j\} = \frac{\exp\{x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}\}}{1 + \sum_{r=1}^J \exp\{x'_{0t}\beta_{0r} + x'_{1rt}\beta_1 + p_{rt}\beta_{ip} + \xi_{rt}\} \kappa_{rt}}. \quad (2)$$

The modification $\kappa_{rt} \equiv \mathbf{1}\{\text{Party } r \text{ runs in District } t\}$ reflects the impact coalition formation has on the menu of parties available to voters in each district. This formulation assumes that voters cast their ballots “sincerely” in favor of their most preferred candidate, without any strategic considerations. While accounting for strategic voting is beyond the scope of this paper, the proportional nature of the post-election allocation of seats and future funding among parties described in Section 3 provides support for the sincere voting assumption.⁶

Let candidate j 's vote share in district t be denoted by s_{jt} . The current setting, in which the expected vote share simply equals the choice probabilities, admits a linear “demand” system. Denoting the share of voters abstaining or writing-in candidates by s_{0t} , the logged vote shares (given a large number of voters) take the form:

$$\mathcal{S}_{jt} \equiv \log s_{jt} - \log s_{0t} = x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}. \quad (3)$$

Among other sources, endogeneity arises from parties' consideration of unobserved local shocks to voter preferences when determining expenditures. This reaction induces correlation between the unobserved shock ξ_{jt} and spending levels p_{jt} . However, these expenditures also respond to L exogenous instruments z_{jt} , allowing us to identify the causal relationship between campaign expenditures and their impact on vote shares. Imposing an (admittedly restrictive) linear structural relationship on expenditures and these features, suppose:

$$p_{jt} = x'_{0t}\pi_{0j} + x'_{1jt}\pi_1 + z'_{jt}\pi_z + \nu_{jt}, \quad \mathbb{E}[\nu_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0. \quad (4)$$

⁶See Kawai and Watanabe (2013) for an example of the challenges involved in identifying strategic voting.

Assumption 1 Linear Logit Truthful Voting Structural Model.

1. In each of T districts, a large and representative sample of voters truthfully vote for their most-preferred candidate under equation (1)'s utility specification.
 2. Logged vote shares are linear in K_0 district characteristics x_{0t} , K_1 candidate characteristics x_{1jt} , and campaign spending p_{jt} according to equation (3).
 3. Campaign spending is linear in the district and candidate characteristics x_{0t} and x_{1jt} as well as L exogenous instruments z_{jt} , as in equation (4).
 4. Residual vote share correlates endogenously with campaign spending, as in equation 5, but is exogenous with respect to instruments: $\mathbb{E}[\eta_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0$.
-

Finally, suppose that district-specific shocks to preferences take a linear form:

$$\xi_{jt} = \rho\nu_{jt} + \eta_{jt}, \quad \mathbb{E}[\eta_{jt}|\nu_{jt}] = 0. \quad (5)$$

We consolidate the above statements about the data generating process for observed vote shares in Assumption (1). This setting presents a simple model for evaluating the influence of campaign spending on voter preferences, in which instrumental variables via two-stage least squares estimation can be used for accurate inference. With standard regularity conditions, consistent inference on β_p can proceed using a standard IV regression of the model:

$$\mathcal{S}_{jt} = x'_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}, \quad \mathbb{E}[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0. \quad (6)$$

Fitting the first stage regression from equation 4 gives fitted values for p_{jt} ,

$$\hat{p}_{jt} = x'_{0t}\hat{\pi}_{0j} + x'_{1jt}\hat{\pi}_1 + z'_{jt}\hat{\pi}_z. \quad (7)$$

The second stage then fits the regression using these fitted values:

$$\mathcal{S}_{jt} = x'_{0t}\hat{\beta}_{0j} + x'_{1jt}\hat{\beta}_1 + \hat{p}_{jt}\hat{\beta}_p + \xi_{jt}, \quad \mathbb{E}[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}, \hat{p}_{jt}] = 0. \quad (8)$$

Under standard conditions, these estimates will be asymptotically normal with the usual

Table 4: Pre-Selected Controls for Fixed Model of Voting

Regional Dummies	Demographics	Economic Status
Region 1	% of Pop Age 18-24	Unemployment
Region 2	% of Pop Age 65+	% of Households w/Car
Region 3	% of Pop that's Married	% of Households w/Refrigerators
Region 4	Average Years of Education	% of Households w/o Basic Utils
Region 5	% of Pop with Elementary Ed	% of Households w/Female Head

This table presents demographic control variables taken from the census measured at the district-level that are included in a pre-specified model of voter preferences. Each of these controls are associated with a party-specific fixed effect, x_{0t} in the utility model.

variance-covariance matrix, admitting standard hypothesis tests for inference.

4.2 A Preselected Model for Returns to Campaign Spending

As a first-pass in our empirical analysis, we present the results for the linear model using a preselected set of control variables adopted by Montero (2015). The selected control variables, summarized in table 4, relate to regional dummies, economic characteristics, education levels, and household structures. We allow for both a linear and quadratic impact of campaign spending on candidate vote share, with the latter reflecting diminishing marginal returns to campaign expenditure. As instruments for campaign spending, we include lagged campaign expenditures, campaign expenditures by competitors in nearby districts, and, as cost shifters, the population density of the district and the percent of the population with internet access.

The main results in table 5's Panel A indicate a positive first-order return to campaign spending, with the linear contribution of campaign expenditure to vote share indicating a 0.63% expected increase from raising campaign spending by \$10,000 USD. The coefficient for squared campaign spending indicates the second-order effect diminishes this contribution by 0.03%. Both of these effects are highly statistically significant.

Investigating the interactive fixed-effects for demographic characteristics shows that many of these fixed-effects are statistically insignificant. Panel B in table 5 reports the p-Value for

Table 5: Estimated Returns to Campaign Spending in Pre-Selected Model (Model LM-F)

Panel A: Main Results					
	Coefficient	Std Error	t-Stat	p-Value	
Expenditures	0.627	0.213	2.95	0%**	
Expenditures ²	-0.033	0.011	3.03	0%**	

Panel B: Significance of Demographic Controls (p-Values)					
Party	Region 1	Region 2	Region 3	Region 4	Region 5
MP	45%	0%**	63%	15%	80%
NA	2%*	12%	29%	51%	92%
PAN	44%	0%**	4%*	1%*	26%
PRI	88%	48%	86%	37%	73%
PVEM	8%	79%	1%**	87%	8%
CM	11%	98%	5%	18%	14%

% of Households with					
Party	Unempl	Car	Refrigerator	Utilities	Female Head
MP	19%	0%**	9%	17%	1%*
NA	26%	41%	0%**	0%**	47%
PAN	39%	13%	1%*	41%	88%
PRI	36%	1%**	23%	9%	3%*
PVEM	52%	8%	59%	50%	35%
CM	11%	81%	63%	64%	0%**

Avg Years					
Party	Pop 18-24	Pop 65+	Married	Element Ed +	School
MP	66%	73%	61%	3%*	1%**
NA	10%	5%*	3%*	9%	8%
PAN	91%	1%**	31%	4%*	29%
PRI	38%	86%	32%	6%	77%
PVEM	2%*	35%	68%	92%	4%*
CM	17%	48%	1%**	37%	49%

Panel A reports the return to campaign expenditures and squared campaign expenditures using the linear logit voting model estimated from equation (8) with interactive fixed effects between the political party and demographic controls listed in Table 4. Panel B reports the significance of each of the interactive fixed-effects by party, with * and ** indicating significance at the 5% and 1% levels, respectively.

the t-Statistics associated with each of these individual fixed effects. The majority (77%) of these coefficients do not have a statistically significant effect on expected vote shares. This result motivates our application of variable selection techniques in the current problem. Can we reduce the number of parameters we need to estimate? Will doing so provide more robust results?

5 Model LM-S: Variable Selection and Inference

Without pre-specifying which demographic controls to include in the analysis, our application includes more potential parameters in the model than we have available observations. These demographic characteristics may be either irrelevant or redundant in describing voting preferences, with many controls reporting similar information with slightly different measures. As such, it's both reasonable and necessary to ignore those demographic variables that have little explanatory power. This variable selection exercise, however, has the potential to distort inference on the effect of campaign spending on vote shares and so our goal is to conduct this exercise while maintaining consistent and valid inference.

As discussed in section 2.1, a considerable literature in econometrics and statistics explores the effect of model selection on inference in cases where the number of variables exceeds the number of observations. Performing inference in this environment depends critically on restricting the data generating process to satisfy some form of “sparsity.” Even though there may be a large number of possible parameters, only a relatively small number of those parameters are truly non-zero in a sparse model. Consequently, estimation requires selecting which variables are actually relevant to the estimation problem and excluding the irrelevant variables. In this subsection, we'll formally introduce the notion of sparsity we assume and review existing results for consistent inference in the simple environment described above.

5.1 Sparsity Assumptions and Regularity Conditions

Sampling approximations for high-dimensional inference require an asymptotic framework that allows the number of parameters (here, $JK_0 + K_1$) to grow with the number of observations. We treat each district as the unit of observation, due to obvious correlation in the vote shares across parties within a given district. For convenience, we'll assume that number of candidate-specific characteristics, K_1 , and the number of excluded instruments, L , remain fixed but allow the number of coefficients associated with district demographic characteristics, K_0 , to become large as $T \rightarrow \infty$. For measurement purposes, introducing sample-size

dependent variable specifications requires indexing the data generating process, which we'll denote \mathcal{P}_T , by the sample size. Within this sequence of data generating processes, we allow $K_T \equiv JK_{0T} + K_1$ to be much larger than sample size T , requiring only $\log(K_T) = o(T^{1/3})$.⁷

Assumption 2 Exact Sparsity in Preferences and Spending

Let $\theta = [\beta'_{01}, \dots, \beta'_{0J}, \beta'_1, \pi'_{01}, \dots, \pi'_{0J}, \pi'_1, \pi'_z]'$. Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^\infty$, has K_T possible parameters. Allowing for the possibility that $K_T > T$, only $k_T < T$ of these parameters are non-zero. We also allow for the possibility that both $K_T \xrightarrow{T \rightarrow \infty} \infty$ and $k_T \xrightarrow{T \rightarrow \infty} \infty$, but we fix the number of excluded instruments in z at $L \geq 2$.

1. The parameter space isn't too large, with $\log(K_T) = o(T^{-1/3})$.
 2. The model is sparse, so the number of non-zero variables $k_T^2 \log^2(K_T \vee T) = o(T^{-1})$
 3. The Gram matrix satisfies a sparse eigenvalue condition that ensures finite-sample identification of the sparse model.
 4. Non-zero coefficients are bounded away from zero.
 5. The distribution of controls, instruments, and vote shares have exponential tails.
-

A more detailed exposition of Assumption (2) appears in Appendix B.2, along with additional commentary. This assumption consolidates the restrictions in Belloni et al. (2013a)'s (ASTE) condition with Belloni et al. (2012)'s (AS) condition in an exactly sparse specification given a fixed number of excluded instruments. The first two conditions ensure that the true model has sufficient structure and enough zeros to be estimable with available data. The third assumption ensures that, even though the complete empirical Gram matrix (" $X'X$ ") will be non-invertible (because $K_T > T$), the sub-matrices formed from the variables associated with non-zero coefficients are almost surely well-behaved. The fourth assumption ensures that non-zero coefficients satisfy the conditions for an oracle inequality to be selected by a Lasso estimator. The last assumption allows the application of large deviation theory to bound the probability of mis-classifying zero-coefficients with a fixed penalty weighting.

Additionally, we require regularity conditions on variability in the data generating process

⁷Note that our limits are taken with respect to a large number of markets ($T \rightarrow \infty$) for a fixed number of products competing in each market. Other analyses of demand data consider the problem where $J \rightarrow \infty$; we maintain J as fixed. We could allow J to grow with the main restriction that $\log(JK_0 + K_1) = o((JT)^{1/3})$, which is already satisfied by assumption 2.1.

Assumption 3 High-Dimensional Linear Logit Regularity Conditions

1. Sufficient moments for unmodeled variability in the data admit a LLN and CLT.
 2. Variability in observables and their impact on unobservables is bounded.
 3. Regularity conditions for asymptotic theory with i.n.i.d. sampling.
 4. Regularity conditions for optimal instruments from first-stage regression.
-

summarized in Assumption (3) for well-behaved asymptotic properties of the post-selection estimator. For completeness, the technical details and additional discussion appears in Appendix B. These conditions present a special case of the regularity conditions embedded in Belloni et al. (2013a) and Belloni et al. (2012)’s condition RF. In their detailed remarks, BCH and BCCH present a number of plausible sufficient conditions that illustrate these assumptions are not overly restrictive. These restrictions are sufficient to apply the asymptotic results for post-selection estimators established in Belloni et al. (2012).

5.2 Selecting Control Variables for Inference

Our inference strategy proceeds in two stages of penalized estimation for selecting control variables followed by a two-state-least-squares estimator using the selected controls in an unpenalized model. The two stages of selection reflect our need to model conditional expectations for the expected impact of control variables on both (I) the campaign spending treatment variable and (II) the vote share outcome. We perform each of the variable selection exercises using a lasso regression, which minimizes the sum of squared residuals subject to an L_1 penalty on the regression coefficients. For ease of reference, we summarize our inference approach in Algorithm (1).

Our variable selection exercise identifies the controls that are important for predicting vote shares and campaign spending. For both of these stages, we apply the iterated Lasso estimator proposed in Belloni et al. (2012) for heteroskedastic, non-Gaussian models. In short, candidate and demographic characteristics that drive variation in p but have little explanatory power for vote shares tend not to be selected by the first lasso (Step I), inviting

Algorithm 1 Post-Selection Estimation and Inference on Treatment Effects after Double-Selection of Controls in High-Dimensional Logit Voting Model

I. **Select controls for expected vote share.** Let $x^I \equiv \{x | \tilde{\beta}_I(x) \neq 0\}$, where:

$$\tilde{\beta}_I = \arg \min_{\beta \in \mathbb{R}^{K_T+1}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_\beta \beta\|_1.$$

II. **Select controls for campaign spending.** Let $x^{II} \equiv \{x | \tilde{\omega}(x) \neq 0\}$, where:

$$\tilde{\omega} = \arg \min_{\omega \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1)^2 + \frac{\lambda_\omega}{T} \|\hat{\Upsilon}_\omega \omega\|_1.$$

III. **Post-Selection Estimation and Inference.** Let $\tilde{x} = x^I \cup x^{II}$ and compute the unpenalized IV regression:

$$\mathcal{S}_{jt} = \tilde{x}'_{0t}\beta_{0j} + \tilde{x}'_{1jt}\beta_1 + p_{jt}\beta_p + \xi_{jt}, \quad \mathbb{E}[\xi_{jt} | p_{jt}, \tilde{x}_{0t}, \tilde{x}_{1jt}, z_{jt}] = 0.$$

Details: $\lambda_{(\beta)} = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2K_T + 2))$ and $\lambda_{(\omega)} = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2K_T))$, with $c = 1.1$ and $\gamma = \frac{0.05}{\log(K_T \vee T)}$, satisfying the restrictions $c > 1$ and $\gamma = o(\log(K_T \vee T))$. $\hat{\Upsilon}_\beta$ is a diagonal matrix whose ideal (k, k) entry is $\sqrt{\bar{E} [x_{k,jt}^2 \epsilon_{jt}^2]}$, with x_k representing the k^{th} regressor and residuals $\epsilon_{jt} = \mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p$. $\hat{\Upsilon}_\omega$ is defined analogously for the regression in step (II). Since the residuals are unobserved, these penalty loadings are feasibly calculated using the iterative algorithm presented in Appendix A.

an omitted variable bias. The second stage of selection (Step II) mitigates this potential distortion by explicitly modeling the campaign spending process.

The Lasso estimator's loss function represents a convex optimization problem with a penalty that enforces many of the estimated coefficients to be exactly zero. As such, it presents a computationally tractable and convenient model selection device. We note here that we allow for the demographic variables selected as relevant interactive fixed-effects to differ across parties. For instance, Unemployment may be a relevant control for the PRI party but not for the other parties. This yields a very flexible control strategy without introducing six new parameters for each demographic variable under consideration.

5.3 Post-Selection Estimation and Inference

The previous subsections focus on penalization strategies for estimating sparse models, particularly as a device for selecting which variables to include and which to exclude from the model. We now take these selected variables and consolidate them into a post-selection model for unpenalized estimation via a control function approach. We collect the controls selected by either of the variable screening devices into $\tilde{x} = x^I \cup x^{II}$, the number of which we denote by \tilde{k}_T . Note that we allow each of the parties to have fixed-effects for different demographic variables and, as such, no longer distinguish between demographic and candidate-specific control variables.

After selection, we can use the usual 2SLS approach to estimate the effect of campaign spending on vote share. Estimating the unpenalized, post-selection, first-stage regression:

$$p_{jt} = \tilde{x}'_{jt}\pi_x + z'_{jt}\pi_z + \tilde{\nu}_{jt}, \text{ gives fitted values, } \tilde{p}_{jt} = \tilde{x}'_{jt}\tilde{\pi}_x + z'_{jt}\tilde{\pi}_z. \quad (9)$$

This first-stage regression gives Belloni et al. (2012)'s optimal instruments in the presence of a high-dimensional vector of controls with a fixed number of instruments. To characterize the residual variation in \tilde{p} after controlling for the selected demographic variables, which will characterize the denominator for our standard errors, compute the regression:

$$\tilde{p}_{jt} = \tilde{x}'_{jt}\psi_x + \check{p}_{jt}, \text{ which gives fitted residuals, } \hat{\check{p}}_{jt} = \tilde{p}_{jt} - \tilde{x}'_{jt}\hat{\psi}_x. \quad (10)$$

Finally, we use the first-stage estimate of exogenous variation in campaign spending and the selected control variables in our second-stage regression.

$$\mathcal{S}_{jt} = \tilde{p}_{jt}\beta_p + \tilde{x}'_{jt}\beta_x + \xi_{jt}, \quad \mathbb{E}[\xi_{jt}|\tilde{x}_{jt}, \tilde{p}_{jt}] = 0. \quad (11)$$

The second-stage results provide a consistent estimator for the treatment effect of campaign spending on candidate vote share. Our regularity conditions imposed thus far are sufficient to yield asymptotic normality of the estimate for β_p , allowing the application of standard

hypothesis tests. This result follows immediately from Theorem 3 in Belloni et al. (2012).

Theorem 1 (Inference on Returns to Campaign Spending under Sparsity). *Suppose Assumptions (1) - (3) hold, then the estimated treatment effect of campaign spending on candidate vote share, $\hat{\beta}_p$, from fitting equation (11) is asymptotically normal:*

$$V_p^{\infty-1/2} \sqrt{JT} \left(\hat{\beta}_p - \beta_p \right) \rightarrow_d N(0, 1), \text{ with } V_p^{\infty} = \bar{E} \left[\check{p}_{jt}^2 \right]^{-2} \bar{E} \left[\check{p}_{jt}^2 \xi_{jt}^2 \right].$$

Here \check{p}_{jt} is the residual exogenous variation in the optimal instrument after regressing \tilde{p}_{jt} on \tilde{x}_{jt} . Letting $\hat{\xi}_{jt} = \mathcal{S}_{jt} - \tilde{p}_{jt} \hat{\beta}_p - \tilde{x}'_{jt} \hat{\beta}_x$ represent the residuals from the regression (11), define:

$$\hat{\Omega} \equiv \frac{1}{JT} \sum_{j,t=1}^{J,T} \hat{p}_{jt}^2 \rightarrow_p \bar{E} \left[\check{p}_{jt}^2 \right]^{-2}, \quad \hat{\Sigma} \equiv \frac{1}{JT} \sum_{j,t=1}^{J,T} \sum_{k,s=1}^{J,T} \rho_{\{jt,ks\}} \hat{p}_{jt} \hat{p}_{ks} \hat{\xi}_{jt} \hat{\xi}_{ks} \rightarrow_p \bar{E} \left[\check{p}_{jt}^2 \xi_{jt}^2 \right],$$

where $\rho_{\{jt,ks\}}$ represents the regularization coefficient for a HAC variance estimator. Consequently, V_p^{∞} can be consistently estimated using $\hat{V}_p^{\infty} \equiv \hat{\Omega}^{-1} \hat{\Sigma} \hat{\Omega}^{-1}$ and replacing $V_p^{\infty-1/2}$ with $\hat{V}_p^{\infty-1/2}$ preserves the t -statistic's asymptotic standard normal distribution.

5.4 Returns to Campaign Expenditures after Control Selection

We now investigate the robustness of our findings on campaign spending from section 4.2 to the fixed specification. Instead of restricting our attention to a fixed set of control variables, we allow for fixed effects driven by any of the 210 demographic variables captured in the Mexican census either linearly, quadratically, or in logs. With six parties' fixed effects, the most unstructured representation of this model allows for over 3,780 parameters when we have only 300 markets and 1,301 total observed market shares.

Table 6's Panel A presents the headline results for the flagship specification with the tuning parameters $c = 1.1$ and $\gamma = 0.05/\log(T \vee K_T)$. This double-lasso model, which includes 20 interactive fixed effects, is much more parsimonious than the pre-selected model, which allowed for 90 such free parameters. With the reduced number of included controls, the magnitude of the effect of \$10,000 in campaign expenditures on vote share declines

Table 6: Campaign Expenditure and Vote Shares with Control Selection (Model LM-S)

Panel A: Main Results									
	Coefficient		Std Error	t-Stat	p-Value				
Expenditures	0.423		0.098	4.32	0%				
Expenditures ²	-0.019		0.006	2.97	0%				

Panel B: Robustness to Tuning Parameters									
Expenditure - Coefficient (Std Err)					Expenditure ² - Coefficient (Std Err)				
$\gamma \log(K_T \vee T)$					$\gamma \log(K_T \vee T)$				
c	0.01	0.05	0.1	0.2	c	0.01	0.05	0.1	0.2
1	0.423 (0.098)	0.421 (0.098)	0.421 (0.098)	0.491 (0.149)	1	-0.019 (0.006)	-0.019 (0.006)	-0.019 (0.006)	-0.027 (0.009)
1.1	0.426 (0.097)	0.423 (0.098)	0.423 (0.098)	0.423 (0.098)	1.1	-0.019 (0.006)	-0.019 (0.006)	-0.019 (0.006)	-0.019 (0.006)
1.25	0.381 (0.098)	0.407 (0.098)	0.398 (0.096)	0.426 (0.097)	1.25	-0.014 (0.007)	-0.016 (0.007)	-0.017 (0.006)	-0.019 (0.006)
1.5	0.378 (0.101)	0.379 (0.098)	0.381 (0.099)	0.381 (0.099)	1.5	-0.014 (0.007)	-0.014 (0.007)	-0.014 (0.007)	-0.014 (0.007)
1.75	0.412 (0.125)	0.387 (0.102)	0.397 (0.102)	0.378 (0.101)	1.75	-0.015 [†] (0.008)	-0.015 (0.007)	-0.015 (0.007)	-0.014 (0.007)
2	0.102 [†] (0.130)	0.182 [†] (0.139)	0.225 [†] (0.119)	0.412 (0.125)	2	0.004 [†] (0.009)	0.000 [†] (0.010)	-0.003 [†] (0.008)	-0.015 [†] (0.008)

[†] Statistically insignificant at the 5% Level

Panel C: Number of Selected Interactive Fixed Effects							
$\gamma \log(K_T \vee T)$	c						
	1.00	1.10	1.25	1.50	1.75	2.00	
0.01	20	19	15	11	8	3	
0.05	21	20	17	12	11	5	
0.1	21	20	18	14	10	6	
0.2	24	20	19	14	11	8	

This table reports the contribution of campaign expenditures to a candidate’s vote share after data-driven selection of demographic controls for interactive fixed-effects using Algorithm 1. Panel A reports the first- and second-order contribution for a benchmark specification with tuning parameters $c = 1.1$ and $\gamma \log(K_T \vee T) = 0.1$. Panel B reports the robustness of this result with respect to these tuning parameters. Panel C indicates the number of interactive fixed-effects included under each of the model specifications.

from 0.63% to 0.42%. However, this reduction in magnitude brings with it much lower standard errors, resulting in a much more significant t-Statistic. The second-order effect is also somewhat reduced, from -0.033% to -0.019%, which remains significant at reasonable confidence levels though slightly less so than in the pre-selected model.

Panel B in Table 6 reflects the sensitivity of these estimates to the tuning parameters

c and γ , illustrating the robustness of these results. Under reasonable specifications of the penalty, the estimated first-order effect ranges from 0.378% to 0.426% with all t-Statistics greater than 3. The second-order effect maintains its significance in these specifications as well, with an effect size ranging from -0.014% to -0.027%. At extreme levels of penalization, the magnitude of the contribution diminishes severely and loses statistical significance in a model that includes only three interactive fixed-effects. These findings are consistent with the guidance by Belloni et al. (2012) that the value for c should be set close to unity and that the selection mechanism may encounter problems when c becomes too large.

Finally, Panel C reports the number of selected control variables under the flagship variable-selection specification. These results illustrate how large penalty terms effectively reduce the model to one in which no controls are included to mitigate observed heterogeneity in voting preferences and campaign expenditures. Interestingly, the number of included controls roughly matches the number of significant controls from the pre-specified model. However, we note that these controls are all different.

6 Model RC-F: Heterogeneous Impressionability with Fixed Controls

The linear vote share model imposes a strong homogeneity restriction on voter preferences within any given district with potentially material empirical implications. In the literature on demand estimation, one approach to allowing for heterogeneity in preferences allows for random-coefficients in the individual's utility model. We incorporate this heterogeneity here by allowing voters to have heterogeneous impressionability, introducing a random-coefficient to the individual influence of campaign spending. Among other sources, these random coefficients could reflect heterogeneous levels of attention paid by different voters in the population.

6.1 GMM Estimation with the BLP Model

We begin by focusing on the structural model for preferences without addressing the variable selection step necessary to address the high dimensionality of the problem. To model heterogeneity in preferences, let an individual voter's preference for candidate j be defined as:

$$u_{ijt} = x_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_{0p} + p_{jt}b_{ip} + \xi_{jt} + \epsilon_{ijt}, b_{ip} \sim N(0, v_p^2). \quad (12)$$

Conditional on b_{ip} , when ϵ_{ijt} has the usual Type-I extreme value distribution, voter i 's decision will be governed by the logit choice probabilities:

$$Pr\{y_{ijt} = j|b_i\} = \frac{\exp\{x_{0t}\beta_{0j} + x'_{1jt}\beta_1 + p_{jt}\beta_p + p_{jt}b_i + \xi_{jt}\}}{1 + \sum_{r=1}^J \exp\{x_{0t}\beta_{0r} + x'_{1rt}\beta_1 + p_{rt}\beta_p + p_{rt}b_i + \xi_{rt}\}}. \quad (13)$$

Since these individual shocks to voter sensitivity aren't observed, we need to integrate equation 13 to compute the expected vote share for a candidate in the district. Letting Φ represent the standard normal distribution's cumulative density, gives:

$$s_{jt} = \int Pr\{y_{ijt} = j|b_i\} d\Phi(b_i/v_p). \quad (14)$$

The nonlinear form of the expected vote share in equation 14 complicates inference because we can no longer transform the vote shares into a generalized linear model. As such, we follow Berry et al. (1995)'s (BLP) approach to identify the model by exploiting the exogeneity of the party-district specific shocks to expected utility. Given a candidate specification for parameter values θ , Berry et al. (1995) show that a contraction mapping recovers these shocks, which we denote $\xi_{jt}(\theta, X, p, s)$. Under the true values for θ , the instruments are orthogonal to these shocks, i.e., $\mathbb{E}[\xi_{jt}(\theta, X, p, s) | z_{jt}] = 0$, so that θ is estimated using a GMM objective function with weighting matrix W :

$$Q(\theta, x, z, p, s) = \frac{1}{JT} \xi_{jt}(\theta, X, p, s)' z W z' \xi_{jt}(\theta, X, p, s). \quad (15)$$

Assumption 4 BLP Truthful Voting Structural Model.

1. In each of T districts, a large number of voters truthfully vote for the candidate they most prefer given the utility specification for u_{ijt} in equation 12.
 2. Expected vote shares are non-linear in K_0 district characteristics x_{0t} , K_1 candidate characteristics x_{1jt} , and campaign spending p_{jt} as in equation 14.
 3. Campaign spending is linear in the district and candidate characteristics x_{0t} and x_{1jt} as well as L exogenous instruments z_{jt} , as in equation 4.
 4. Residual shocks to expected voter preferences in a district are endogenously correlated with unmodeled variation in campaign spending, as in equation 5, but has a zero conditional expectation given district and candidate district characteristics, observable instruments: $\mathbb{E}[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0$.
-

In the standard setting with a fixed number of controls and instruments, for any positive-definite W , minimizing equation 15 provides an asymptotically normal estimator for the parameters in θ . To address numerical issues in the evaluation of this estimator, Dube et al. (2012) present an MPEC algorithm, which we also use here.

One last sensitivity associated with the GMM objective function above relates to the instruments themselves. Berry et al. (1999) present an early discussion on the importance of using Chamberlain (1987) optimal instruments in evaluating (15). Gandhi and Houde (2015) illustrate how to utilize vote shares themselves as valuable instruments. Reynaert and Verboven (2014) illustrate how sensitive the estimator is to implementation with optimal instruments, particularly with respect to estimating the variance parameters v_p . Our implementation adopts this latter approach, since the Reynaert and Verboven (2014) instruments are easily recovered from the gradient of the constraints in the MPEC algorithm.

Assumption (5) contains regularity conditions, which are fairly standard for GMM estimation with i.n.i.d. data (additional technical details for these conditions are presented in Appendix B). By textbook analysis, assumptions (4) and (5) are sufficient to establish the usual consistency and asymptotic normality results for the value of θ that minimizes equation (15).

Assumption 5 Regularity Conditions for GMM Estimator

1. Compactness of parameter space: The true parameter values $\theta_0 \in \Theta_{KT}$, where $\Theta_K \subset \mathbb{R}^{KT+2}$ is compact, with a compact limit set $\Theta_\infty \equiv \lim_{T \rightarrow \infty} \Theta_K$.
 2. Continuity and differentiability of sample-analog and population moment conditions in parameter space.
 3. Letting $g_{jt}(\theta) \equiv [x'_{0t}, x'_{1jt}, z'_{jt}]' \xi_{jt}(\theta)$, a uniform law of large numbers ensures the sample-analog, $\frac{1}{JT} \sum_{j,t=1}^{JT} g_{jt}(\theta)$, converges to the population moment condition.
 4. A uniform law of large numbers applies to Hessian of sample analog to the population moment condition, $\hat{G}_T(\theta) \equiv \frac{1}{JT} \sum_{j,t=1}^{JT} \frac{\partial g_{jt}(\theta)}{\partial \theta'}$.
 5. The weighting matrix, W_T , is positive definite and converges to W , a symmetric, positive definite, and finite matrix.
 6. The expected outer product of the score, $\Omega \equiv \lim_{T \rightarrow \infty} (JT)^{-1} \sum_{j,t=1}^{JT} \mathbb{E}[g_{jt}(\theta) g_{jt}(\theta)']$, is a positive definite, finite matrix.
 7. The matrix $\Sigma \equiv G(\theta_0)' \Omega^{-1} G(\theta_0)$ is almost surely positive definite and finite.
-

6.2 Campaign Spending with Heterogeneous Impressionability

We again begin our empirical analysis of heterogeneous impressionability in voting for Mexico using the pre-specified set of controls considered in table 4. These results allow us to differentiate the influence of heterogeneity in the model from the effects of control selection.

Table 6.2's Panel A reports the expected coefficients and standard deviation of coefficients associated with campaign expenditures' influence on voters' latent utility. The results indicate that heterogeneous impressionability is not a prominent feature of preferences, as revealed through the low variance of the coefficients themselves, which are not statistically distinguishable from zero.

Partly as a consequence of this limited heterogeneity, the expected coefficients are rather similar to those reported in Table 6. Estimating standard errors with the sandwich covariance matrix we find weaker, but still significant, evidence for the significance of campaign expenditure on candidate vote share. The linear term's p-value rises to 4% and the negative quadratic effect loses statistical significance with a p-value of 17%.

As in our analysis of the linear model, 6.2’s Panel B reports the significance of the demographic controls included in the model with heterogeneous impressionability. With the higher variance of the estimates under the random coefficients specification, we find a slightly smaller share of the interactive fixed effects achieve the threshold for a statistically significant influence on voter preferences.

7 Model RC-S: Variable Selection in the BLP Voting Model

We now address the implications of the high-dimensional setting on the BLP model, particularly when there are more control variables than observations. Inference in this setting is non-trivial, since the model is unidentified in finite samples. That is, there exists a multiplicity of values for the parameters θ for which the residual shock to preferences, ξ_{jt} , can be equal to zero for all observations. As in the linear setting, a sparsity condition on the coefficients suggests incorporating a penalty to the criterion function that admits consistent inference. Our approach builds on the proposed technique from Gillen et al. (2014), which presents a model for inference in demand models after selection from a high-dimensional set of product characteristics when the number of control variables is of the same order of magnitude as the number of markets.

Our analysis here extends the Gillen et al. (2014) approach to a “non-polynomial” setting that allows the number of possible control variables to grow exponentially with the number of markets. Implementing selection in this “ultra-high” dimensional setting must confront two significant complications. The first is computational, as optimizing nonlinear objective functions with a non-polynomial number of parameters is simply infeasible in most circumstances. The second is analytical, in that we must apply the oracle properties established in Fan and Liao (2014)’s analysis of penalized estimation in high-dimensional GMM problems. We begin this section by discussing the conditions for valid inference under a penalized GMM objective function before turning to the computational issues and empirical results.

Table 7: Campaign Expenditure and Vote Shares with Heterogeneous Impressionability (Model RC-F)

Panel A: Main Results					
Expected Coefficients	Coefficient	Std Error	t-Stat	p-Value	
Expenditures	0.77	0.38	2.04	4%	
Expenditures ²	-0.04	0.03	-1.39	17%	
Variance of Coefficients	Coefficient	Std Error	t-Stat	p-Value	
Expenditures	0.11	0.40	0.27	79%	
Expenditures ²	0.00	0.04	0.00	100%	

Panel B: Significance of Demographic Controls (p-Values)					
Party	Region 1	Region 2	Region 3	Region 4	Region 5
MP	15%	0%*	84%	35%	85%
NA	2%*	11%	52%	41%	53%
PAN	0%*	0%*	4%*	10%	47%
PRI	62%	46%	89%	28%	58%
PVEM	11%	43%	1%*	95%	10%
CXM	32%	98%	10%	29%	27%

% of Households with					
Party	Unempl	Car	Refrigerator	Utilities	Female Head
MP	29%	0%*	14%	18%	5%
NA	24%	51%	0%*	0%*	71%
PAN	50%	30%	0%*	44%	96%
PRI	40%	2%*	22%	10%	5%*
PVEM	35%	16%	35%	58%	44%
CXM	28%	90%	64%	86%	4%*

Avg Years					
Party	Pop 18-24	Pop 65+	Married	Element Ed +	School
MP	81%	92%	50%	13%	2%*
NA	6%	2%*	2%*	12%	7%
PAN	95%	2%*	23%	6%	46%
PRI	36%	99%	35%	13%	77%
PVEM	1%*	30%	65%	70%	4%*
CXM	38%	52%	10%	58%	50%

Panel A reports the return to campaign expenditures and squared campaign expenditures using the nonlinear BLP voting model with heterogeneous impressionability estimated from equation (15) with interactive fixed effects between the political party and demographic controls listed in Table 4. Panel B reports the significance of each of the interactive fixed-effects by party, with * and ** indicating significance at the 5% and 1% levels, respectively.

7.1 Sparsity Assumptions and Regularity Conditions

We begin by laying out the sparsity conditions required for establishing the oracle properties from Fan and Liao (2014)'s analysis of penalized estimation in high-dimensional GMM prob-

Assumption 6 Sparsity Assumptions for High-Dimensional BLP Model

Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^\infty$, has $K_T > T$ possible parameters, $k_T < T$ of which are non-zero, where both $K_T \xrightarrow{T \rightarrow \infty} \infty$ and $k_T \xrightarrow{T \rightarrow \infty} \infty$. Further, the number of excluded instruments in z is fixed at $L \geq 2$.

1. The parameter space isn't too large, with $\log(K_T) = o(T^{-1/3})$.
 2. The model is sparse, with the number of non-zero variables, $k_T^3 \log k_T = o(T^{-1})$
 3. The Gram matrix for controls with non-zero influence on vote shares is almost surely positive definite with finite eigenvalues.
 4. The Hessian of the objective function with respect to non-zero variables is almost surely positive definite.
 5. Non-zero coefficients are bounded away from zero.
 6. The marginal distributions for controls, instruments, and residual vote shares have exponentially decaying tails.
-

lems. These properties allow us to generalize the results from Gillen et al. (2014) to apply to a higher-dimensional setting. Gillen et al. (2014) relied on the asymptotic theory of Caner and Zhang (2013), who establish oracle properties for Zou and Hastie (2005)'s Elastic Net in an environment where the number of parameters grows more slowly than the number of observations, so that $K_T/T \rightarrow 0$. In contrast, the oracle properties in Fan and Liao (2014) apply to the setting where the dimensionality grows non-polynomially with the sample size, requiring only that $\log(K_T)/T^{1/3} \rightarrow 0$. These requirements are summarized in assumption (6), which slightly tightens the restrictions of the linear model in assumption (2), with the technical details relegated to Appendix B.

These assumptions are sufficient to establish an oracle property for the lasso-penalized GMM estimator, ensuring that the penalized estimator accurately identifies all non-zero coefficients and effectively thresholds all irrelevant coefficients at zero. Coupled with the continuity and uniform laws of large numbers of the GMM objective function from assumption 5, the results from Fan and Liao (2014) establish sufficient conditions on the penalty term for the penalized GMM estimator to achieve the near oracle convergence rate.

The first stage of inference requires solving a penalized GMM objective function with a lasso penalty. Similar to the linear demand model, we apply a data-dependent penalization

that is robust to heteroskedasticity in sampling across markets:

$$\tilde{\theta} = \arg \min_{\theta \in \mathbb{R}^{K_T+2}} Q(\theta, x, z, p, s) + \frac{\lambda_\theta}{T} \|\hat{\Upsilon}_\theta \theta\|_1. \quad (16)$$

Our penalty loading, $\lambda_\theta = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2(K_T + 4)))$ satisfies the restrictions in Fan and Liao (2014) for a lasso penalty loading to be $k_T\sqrt{k_T}/\sqrt{T} \prec \lambda_\theta/T \prec 1/\sqrt{k_T}$ when $c = 0.05$ and $\gamma = 0.1/\log(K_T \vee T)$. As in the linear model, Υ_θ is a diagonal matrix, with the ideal weights for $\beta_{0j,k}$ equal to $\sqrt{\bar{E}[x_{0t,k}^2 \xi_{jt}^2]}$, for $\beta_{1,k}$ equal to $\sqrt{\bar{E}[x_{1jt,k}^2 \xi_{jt}^2]}$, and for β_p equal to $\sqrt{\bar{E}[p_{jt}^2 \xi_{jt}^2]}$. For the heterogeneity coefficient, v_p , the ideal value in the Υ_θ matrix is $\sqrt{\bar{E}\left[\frac{\partial \xi_{jt}(\theta, x, z, p, s)}{\partial v_p}^2 \xi_{jt}^2\right]}$. Since ξ_{jt} is unobserved, Appendix A reports the feasible iterated algorithm used to calculate Υ_θ .

7.2 Implementing Variable Selection via Penalized GMM

We now describe the approach we use for selecting variables using a penalized GMM estimator, since it is computationally infeasible to directly optimize the GMM objective function in extremely high-dimensional problems. We begin by fitting the nonlinear model with heterogeneous impressionability to the model selected by Algorithm 1. Within this fitted model, we can approximate the optimal instruments for the nonlinear features of the model, selecting the relevant demographic controls for observed heterogeneity in these features. We augment the selected demographic controls from Algorithm 1 with the controls for the latent utilities and optimal instruments for nonlinear features of the model. With this robustly augmented set of control variables, we compute unpenalized GMM estimates for the selected variables. We then verify that the first order conditions for the selected model are satisfied in the larger model with all included controls. The steps presented in algorithm 2 summarize our approach to this implementation.

Our implementation picks up from where Algorithm 1 leaves off, defining the post-selection model including demographic controls \tilde{x} . With this model, we solve the GMM

Algorithm 2 Post-Selection Estimation and Inference with Double-Selection from High-Dimensional Controls in a Voting Model with Heterogeneous Impressionability

- I. **Apply Algorithm 1** to select $\tilde{x} = x^I \cup x^{II}$ as the controls for a homogeneous model.
 II. **Compute GMM estimates for heterogeneous model using selected controls.**

Let $\tilde{\delta}_{jt} \equiv \tilde{x}'_{jt}\tilde{\beta}_j + x'_{1jt}\tilde{\beta}_1 + p_{jt}\tilde{\beta}_p + \xi_{jt}(\tilde{\theta}, x, z, p, s)$, where $\tilde{\theta} \equiv [\tilde{\beta}'_1, \dots, \tilde{\beta}'_J, \tilde{\beta}_1, \tilde{v}_p]'$:

$$\tilde{\theta} = \arg \min Q(\theta, \tilde{x}, z, p, s).$$

- III. **Estimate optimal instruments for heterogeneity in impressionability.** Compute the derivative of the moment condition with respect to the variability parameter v_p :

$$\tilde{z}_{v,jt} = \frac{\partial}{\partial v_p} \xi_{jt}(\theta, \tilde{x}, z, p, s) |_{\theta=\tilde{\theta}}.$$

- IV. **Select controls for mean utilities.** Let $x^{III} \equiv \{x | \tilde{\phi}(x) \neq 0\}$, where:

$$\tilde{\phi} = \arg \min_{\phi \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1)^2 + \frac{\lambda_\phi}{T} \|\hat{Y}_\phi \phi\|_1.$$

- V. **Select controls for optimal nonlinear instruments.** Let $x^{IV} \equiv \{x | \tilde{\zeta}(x) \neq 0\}$, where:

$$\tilde{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\tilde{z}_{v,jt} - x'_{0t}\zeta_{0j} - x'_{1jt}\zeta_1)^2 + \frac{\lambda_\zeta}{T} \|\hat{Y}_\zeta \zeta\|_1.$$

- VI. **Post-selection estimation and inference.** Let $\tilde{x}^* = \tilde{x} \cup x^{III} \cup x^{IV}$ and compute the unpenalized GMM estimate:

$$\tilde{\theta}^* = \arg \min Q(\theta, \tilde{x}^*, z, p, s).$$

- VII. **Verify First Order Conditions in Unselected Model.** For each excluded demographic control x_{0k} , define $\tilde{x}^k = \tilde{x}^* \cup x_{0k}$. Verify the first order improvement in the objective function from including this variable x_{0k} for any party is dominated by the penalty:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q(\tilde{\theta}^*, \tilde{x}^k, z, p, s) < \lambda_\theta \Upsilon_{\theta, (k,k)}, k = 1, \dots, K_0, j = 1, \dots, J.$$

- VIII. **Add improperly excluded variables to the model and iterate.** Define the set of controls that fail to satisfy first order conditions in step (VII) as $x^V = \{x_k : q_k > \lambda_\theta \Upsilon_{\theta, (k,k)}\}$. Redefine $\tilde{x} = \tilde{x}^* \cup x^V$ and return to Step (II) until there are no changes in the set of included variables.

Details: $\lambda_\phi = \lambda_\zeta = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2K_T))$, and $\lambda_\theta = 2c\sqrt{T}\Phi^{-1}(1 - \gamma/(2K_T + 8))$ with $c = 1.1$ and $\gamma = \frac{0.05}{\log(K_T \sqrt{T})}$. The details for calculating the diagonal factor loading matrices $\hat{Y}_{(\cdot)}$, whose ideal entries reflect the square root of the expected product of the squared residual and control variable, are discussed in the text. The iterative algorithms by which we feasibly calculate these values are detailed in Appendix A.

objective function without any penalization:

$$\tilde{\theta} = \arg \min Q(\theta, \tilde{x}, z, p, s). \quad (17)$$

Given the solution $\tilde{\theta}$, we can recover the latent mean utilities:

$$\tilde{\delta}_{jt} = \tilde{x}'_{jt}\tilde{\beta}_j + x'_{1jt}\tilde{\beta}_1 + p_{jt}\tilde{\beta}_p + \tilde{\xi}_{jt}.$$

These provide the outcome variable for which we need to select the relevant demographic controls using another application of the lasso. In parallel to the linear model's treatment, let $x^{III} \equiv \{x | \tilde{\phi}(x) \neq 0\}$, where:

$$\tilde{\phi} = \arg \min_{\phi \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J \left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_{1j} \right)^2 + \frac{\lambda_\phi}{T} \|\hat{\Upsilon}_\phi \phi\|_1. \quad (18)$$

The penalization term λ_ϕ has the same expected form as previous applications. The Υ_ϕ matrix requires a slight adjustment to account for estimation error in the $\tilde{\delta}_{jt}$'s. Defining $\epsilon_{\delta,jt} \equiv \delta_{jt} - \tilde{\delta}_{jt}$ and $\epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_{1j}$, the ideal weight for $\zeta_{0j,k}$ is equal to $\sqrt{\bar{E} [x_{0t,k}^2 (\epsilon_{\delta,jt} + \epsilon_{\phi,jt})^2]}$ and $\sqrt{\bar{E} [x_{1jt,k}^2 (\epsilon_{\delta,jt} + \epsilon_{\phi,jt})^2]}$ for $\beta_{1,k}$. The additional residuals can be characterized by using the asymptotic covariance matrix for $\tilde{\theta}$, which can be consistently estimated using the sandwich covariance matrix from the penalized GMM estimator.

We note that, by applying Algorithm 1, we have already selected the demographic controls necessary to explain observable variation in campaign expenditure. Now we select the demographic controls that explain variation across districts in the heterogeneity of impressionability. To do this, we need the optimal instruments for the heterogeneity parameters to identify the relevant controls for their first-order impact on model fit. Using the fitted model from equation (17), we compute the derivative of the objective function with respect to v_p :

$$\tilde{z}_{v,jt} = \frac{\partial}{\partial v_p} \xi_{jt}(\theta, \tilde{x}, z, p, s) |_{\theta=\tilde{\theta}}.$$

The formula for $\tilde{z}_{v,jt}$ from Berry et al. (1999) is presented in Nevo (2000)’s appendix and easily recovered from the Jacobian of the constraint for the MPEC objective function. We can then select demographic control variables that explain heterogeneity in this optimal instrument using a last application of the lasso estimator. Let $x^{IV} \equiv \{x | \tilde{\zeta}(x) \neq 0\}$, where:

$$\tilde{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\tilde{z}_{v,jt} - x'_{0t} \zeta_{0j} - x'_{1jt} \zeta_1)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1. \quad (19)$$

Since $\tilde{z}_{v,jt}$ represents a generated regressor, we may wish to incorporate the variance induced by estimation error in its definition when determining the adapted penalty factor for equation (19), as in Υ_ϕ . However, by defining $\tilde{z}_{v,jt}$ as our identifying instrument, we only need to select demographic controls for variation in the *generated* $\tilde{z}_{v,jt}$ without regard to the population $z_{v,jt}$, which plays no direct role in our estimation. Consequently, when computing the values for Υ_ζ , we ignore the generated-regressors problem.

7.3 Post-Selection Inference via Unpenalized GMM

Combining the selected controls from Algorithm 1, \tilde{x} with x^{III} and x^{IV} , define $\tilde{x}^* = \tilde{x} \cup x^{III} \cup x^{IV}$. We then compute the unpenalized, post-selection estimator

$$\tilde{\theta}^* = \arg \min Q(\theta, \tilde{x}^*, z, p, s). \quad (20)$$

To maximize the efficiency of our estimates, we first compute the optimal instruments for the Berry et al. (1995) model as discussed in Berry et al. (1999) and Reynaert and Verboven (2014). For the demographic controls and candidate characteristics, the selected variables themselves present the optimal instruments. We computed the optimal instruments for heterogeneity, \tilde{z}_v , in the variable selection stage. Finally, the optimal instruments for campaign expenditures can be easily estimated by an unpenalized first-stage regression which contains

the selected controls and excluded instruments as regressors:

$$\tilde{z}_{p,jt} = \tilde{x}_{jt}^* \hat{\pi}_x + z_{jt}' \hat{\pi}_z.$$

Denoting the optimal instruments by \tilde{z} and the selected control variables by \tilde{x} , we then compute the post-selection estimator for the voting model with heterogeneous impressionability as the solution to:

$$\theta = \arg \min_{\theta \in \mathbb{R}^{k_T+2}} Q(\theta, \tilde{x}, \tilde{z}, p, s). \quad (21)$$

The last step then verifies that this solution also satisfies the first-order conditions for the penalized objective function (16) to ensure we have not erroneously excluded any variables. We perform this test sequentially, evaluating the first-order conditions with respect to each excluded variable and verifying that they are dominated by the magnitude of the penalty term. As when calculating the optimal instruments for the variance parameters in the model, these gradients can be recovered from the Jacobian of the constraint in the MPEC objective function:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q(\tilde{\theta}^*, \tilde{x}^k, z, p, s) < \lambda_\theta v_k, k = 1, \dots, K_0, j = 1, \dots, J.$$

Any variables whose first-order conditions dominate the penalty should be included within the selected model. This requirement leads to an iterative process that, in our experience, converges within two iterations.

Given assumptions 5 and 6, the post-selection estimator for the treatment effects β_p and their heterogeneity v_p will be consistent and asymptotically normal. It is straightforward to show that the oracle property established in Fan and Liao (2014) satisfies the High-Level Model Selection condition in Belloni et al. (2012), giving the asymptotic result:

Theorem 2 (Inference on Returns to Campaign Spending under Sparsity). *Suppose Assumptions (4) - (6) hold, then the estimated effect of campaign spending on mean latent*

voter utilities is asymptotically normal. That is:

$$V_p^{\infty-1/2} \sqrt{JT} \left(\hat{\beta}_p - \beta_p \right) \rightarrow_d N(0, 1).$$

The asymptotic variance V_p^∞ is the element in the sandwich covariance matrix for $\tilde{\theta}^*$ corresponding to β_p . Specifically, defining the limit of the i.n.i.d. expectations:

$$G_0 \equiv \bar{E} \left[\frac{\partial g_{jt}(\theta)}{\partial \theta'} \right] \Big|_{\theta=\theta_0}, \quad \Omega_0 \equiv \bar{E} [g_{jt}(\theta_0) g_{jt}(\theta_0)'], \quad \text{and} \quad \Sigma_0 \equiv G(\theta_0)' \Omega^{-1} G(\theta_0),$$

then $V_\theta^\infty \equiv (G_0' G_0)^{-1} \Sigma_0 (G_0' G_0)^{-1}$, which can be estimated with:

$$\hat{G}_T \equiv \frac{1}{JT} \sum_{j,t=1}^{J,T} \frac{\partial g_{jt}(\theta)}{\partial \theta'} \Big|_{\theta=\tilde{\theta}^*} \rightarrow_p G_0, \quad \text{and} \quad \hat{\Omega} \equiv \frac{1}{JT} \sum_{j,t=1}^{J,T} g_{jt}(\tilde{\theta}^*) g_{jt}(\tilde{\theta}^*)' \rightarrow_p \Omega_0.$$

Consequently, $\hat{\Sigma}_T \equiv \hat{G}_T' \hat{\Omega} \hat{G}_T \rightarrow_p \Sigma_0$ and

$$\hat{V}_\theta^\infty \equiv \left(\hat{G}_T' \hat{G}_T \right)^{-1} \hat{\Sigma}_T \left(\hat{G}_T' \hat{G}_T \right)^{-1} \rightarrow V_\theta^\infty,$$

so replacing $V_p^{\infty-1/2}$ with $\hat{V}_p^{\infty-1/2}$ preserves the t -statistic's asymptotic normal distribution.

Further, if $v_p > \delta > 0$, the estimated variance of campaign spending's impact on voter utilities is also asymptotically normal with asymptotic variance of the estimate given by the sandwich covariance matrix.

7.4 Heterogeneous Impressionability after Control Selection

Even with only validating local optimality conditions for the selected model, practical implementation of Algorithm 2 is quite computationally intensive. In analyzing the Mexican voting data, fitting a single penalty specification requires approximately 80 core-hours of computation. Due to the intensive computational resources required for estimation, we focus our empirical analysis on the benchmark penalty specification where $c = 1.1$ and $\gamma \log(K_T \vee T) = 0.10$.

Panel A of Table 8 reports the impact of campaign expenditures on mean utilities. The

Table 8: Campaign Expenditures and Candidate Vote Share with Heterogeneous Impressionability and Selected Controls (Model RC-S)

Panel A: Main Results				
Expected Coefficients	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.465	0.198	2.35	2%
Expenditures ²	-0.020	0.015	-1.33	18%
Variance of Coefficients	Coefficient	Std Error	t-Stat	p-Value
Expenditures	0.05	0.67	0.07	94%
Expenditures ²	0.00	0.06	0.01	99%

Panel B: Selected Demographic Controls		
MP	NA	PAN
Party FE (+)**	Party FE (-)	Party FE (-)**
Female Popn >60 (-)**	% Female HoH (-)	
	% Pop 18-24 (-)	
	% Pop Married (+)	
PVEM	% Pop w/Elementary Ed (-)*	CXM
Party FE (-)**	Female Popn >60 (-)	Party FE (-)**
% HH w/o Utils (-)	Total Popn >65 (-)	Region 4 Dummy (-)
% HH w/o Fridge (-)	Male Pop >15 w/Ed (-)*	% Female HoH (-)**
Total Popn >65 (-)	Popn in Private Dwell (+)**	Female Pop w/o Ed (+)**
Popn in Private Dwell (+)**	Avg Num People/Dwell (-)	Pop w/Social Medicine (+)
Avg Num People/Dwell (-)**		Travel Time to Polls (+)
	PRI	
	Party FE (-)**	

estimated first-order mean effect of 0.465 is a bit lower than the pre-selected model's result of 0.767. This result matches the impact of variable selection on the estimated effect in the linear model, which dropped to 0.423 from 0.627 after variable selection. The standard error of the model with variable selection is lower than the pre-specified model's, yielding very similar t-Statistics for both the first- and second-order effects. As in the pre-selected model, the variance coefficients reflecting heterogeneity in preferences are indistinguishable from zero, indicating that there may not be much heterogeneity in voter impressionability.

Panel B of Table 8 reports the actual demographic controls affecting voters' preferences for each party's candidates. For the two largest parties, PAN and PRI, only the party fixed effects were selected to control for voter preferences. For the coalition parties, more controls

were incorporated to reflect heterogeneity in preferences. Interestingly, these demographic controls were most important for characterizing voter preferences for parties that sit in the middle of the policy spectrum. Panel B also indicates the significance and sign of the controls' impact on voter preferences. However, we caution against drawing too many conclusions from their selection. These controls were selected to identify the impact of campaign expenditures on voter preferences, not as independent causal elements driving voter preferences. As such, their selection may only be as proxies representing the effect of other variables on preferences and spending.

8 Generalizations and Extensions

Our results imposed an exact sparsity condition, requiring that the negligible coefficients in the model are exactly equal to zero. This restriction is quite a bit stronger than the approximate sparsity condition presented in Belloni et al. (2012) and Belloni et al. (2013a). The sampling properties of the penalized GMM estimator presented in Fan and Liao (2014) are robust to local perturbations, suggesting that exact sparsity could be weakened to the generalized forms of sparsity considered by Zhang and Huang (2008) and Horowitz and Huang (2010). Some additional regularity conditions may be required on the nonlinear features recovering latent mean utilities from observed market shares, but such an extension should be viable.

Though our analysis focused on GMM as an approach for estimating the voting model, alternative estimation strategies could also be considered after performing a selection step. Empirical likelihood approaches following Kitamura (2001) and Donald et al. (2003) have been adapted to demand estimation by Moon et al. (2014) and Conlon (2013). The Fan and Liao (2014) asymptotic analysis also applies to empirical likelihood estimators that could be mapped onto these techniques.

Though we allow for correlation among vote shares across parties within a district, we note that our analysis leans heavily on an independence assumption for sampling across districts.

Limited spatial correlation could be accounted for by computing robust standard errors in estimating the covariance matrix of residuals. As long as strong-mixing and ergodicity conditions are met, this sort of dependence should not preclude effective variable selection.

There is some tension between our assumption of a linear campaign financing rule in light of a structural model of competition between parties. Indeed, [Montero \(2015\)](#) solves the equilibrium campaign financing rule in the Mexican election environment and shows it to be highly nonlinear. One way to address this issue characterizes the linearized campaign finance rule as an approximation to the structural finance rule, bounding the approximation error relative to instrumental variability, and showing that the approximation error doesn't affect variable selection and inference. Another strategy might adopt a control function approach to estimation, perhaps following [Kawai \(2014\)](#)'s strategy of incorporating techniques from production function estimation.

9 Conclusion

We present several results in high-dimensional inference and apply these techniques in an empirical analysis of voting behavior in Mexican elections. Our analysis applies high-dimensional inference techniques for estimating aggregate demand models with a very large number of demographic covariates. Though our statistical analysis is largely informed by previously established properties of these techniques, the extensions to the specific application are not trivial.

Our results show, robustly, that campaign expenditures have a significant and positive impact on voters' latent utilities for a candidate, with indications that the impact of these expenditures diminishes with the amount of campaign spending. Strikingly, we find little evidence of heterogeneity in voters' response to campaign expenditure, perhaps because limited variability in the slate of candidates provides little opportunity for this heterogeneity to impact vote shares.

References

- Donald WK Andrews. Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 67(3):543–563, 1999.
- Donald WK Andrews and Biao Lu. Consistent model and moment selection procedures for gmm estimation with application to dynamic panel data models. *Journal of Econometrics*, 101(1):123–164, 2001.
- Alexandre Belloni, Victor Chernozhukov, and Lie Wang. Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4):791–806, 2011.
- Alexandre Belloni, Daniel Chen, Victor Chernozhukov, and Christian Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429, 2012.
- Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies*, 2013a.
- Alexandre Belloni, Victor Chernozhukov, and Ying Wei. Honest confidence regions for logistic regression with a large number of controls. *arXiv preprint arXiv:1304.3969*, 2013b.
- Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890, 1995.
- Steven Berry, James Levinsohn, and Ariel Pakes. Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review*, 89(3):400–430, 1999.
- Peter J Bickel, Ya'acov Ritov, and Alexandre B Tsybakov. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, pages 1705–1732, 2009.
- Jelena Bradic, Jianqing Fan, and Weiwei Wang. Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):325–349, 2011.
- Emmanuel Candes and Terence Tao. The dantzig selector: statistical estimation when p is much larger than n . *The Annals of Statistics*, pages 2313–2351, 2007.
- Mehmet Caner. Lasso-type gmm estimator. *Econometric Theory*, 25(01):270–290, 2009.
- Mehmet Caner and Hao Helen Zhang. Adaptive elastic net for generalized methods of moments. *Journal of Business & Economic Statistics*, (just-accepted), 2013.
- Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987.

- Hai Che, Ganesh Iyer, and Ravi Shanmugam. Negative advertising and voter choice. Technical report, working paper, University of Southern California, 2007.
- Xu Cheng and Zhipeng Liao. Select the valid and relevant moments: An information-based lasso for gmm with many moments. *Journal of Econometrics*, 186(2):443–464, 2015.
- Christopher T Conlon. The empirical likelihood mpec approach to demand estimation. *Available at SSRN*, 2013.
- Consulta Mitofsky. Geometría electoral en México (electoral geometry in Mexico). Public Opinion Poll, http://www.consulta.mx/web/images/MexicoOpina/2013/NA_GEOMETRIA_ELECTORAL.pdf (in Spanish), 2012.
- Arianna Degan and Antonio Merlo. A structural model of turnout and voting in multiple elections. *Journal of the European Economic Association*, 9(2):209–245, 2011.
- Stephen G Donald, Guido W Imbens, and Whitney K Newey. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117(1):55–93, 2003.
- Jean-Pierre Dube, Jeremy Fox, and Che-Lin Su. Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation. *Econometrica*, 2012. forthcoming.
- Robert S Erikson and Thomas R Palfrey. Equilibria in campaign spending games: Theory and data. *American Political Science Review*, 94(03):595–609, 2000.
- Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.
- Jianqing Fan and Yuan Liao. Endogeneity in high dimensions. *Annals of statistics*, 42(3):872, 2014.
- Jianqing Fan and Jinchi Lv. Nonconcave penalized likelihood with np-dimensionality. *Information Theory, IEEE Transactions on*, 57(8):5467–5484, 2011.
- A. Gandhi and J.F. Houde. Measuring substitution patterns in differentiated products industries the missing instruments. *Mimeo*, 2015.
- Eric Gautier and Alexandre Tsybakov. High-dimensional instrumental variables regression and confidence sets. *arXiv preprint arXiv:1105.2454*, 2011.
- Alan Gerber. Estimating the effect of campaign spending on senate election outcomes using instrumental variables. *American Political Science Review*, 92(02):401–411, 1998.

- Alan Gerber. Field experiments in political science. *Handbook of Experimental Political Science*, pages 115–40, 2011.
- Benjamin J Gillen, Hyungsik Roger Moon, and Matthew Shum. Demand estimation with high-dimensional product characteristics. *Bayesian Model Comparison*, pages 301–24, 2014.
- Brett R Gordon and Wesley R Hartmann. Advertising effects in presidential elections. *Marketing Science*, 32(1):19–35, 2013.
- Brett R Gordon, Mitchell J Lovett, Ron Shachar, Kevin Arceneaux, Sridhar Moorthy, Michael Peress, Akshay Rao, Subrata Sen, David Soberman, and Oleg Urminsky. Marketing and politics: Models, behavior, and policy implications. *Marketing Letters*, 23(2): 391–403, 2012.
- Donald Philip Green and Jonathan S Krasno. Salvation for the spendthrift incumbent: Reestimating the effects of campaign spending in house elections. *American Journal of Political Science*, pages 884–907, 1988.
- Matthew Harding and Carlos Lamarche. Penalized quantile regression with semiparametric correlated effects: An application with heterogeneous preferences. *Mimeo*, 2015.
- Joel L Horowitz and Jian Huang. The adaptive lasso under a generalized sparsity condition. *Manuscript, Northwestern University*, 2010.
- Jian Huang, Joel L Horowitz, and Shuangge Ma. Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics*, 36(2):587–613, 2008.
- Gary C Jacobson. The effects of campaign spending in congressional elections. *American Political Science Review*, 72(02):469–491, 1978.
- Gary C Jacobson. Money and votes reconsidered: Congressional elections, 1972–1982. *Public choice*, 47(1):7–62, 1985.
- Kei Kawai. Campaign finance in us house elections. *Mimeo*, 2014.
- Kei Kawai and Yasutora Watanabe. Inferring strategic voting. *The American Economic Review*, 103(2):624–662, 2013.
- Gary King. A solution to the ecological inference problem, 1997.
- Yuichi Kitamura. Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, 69(6):1661–1672, 2001.
- Hannes Leeb and Benedikt M Pötscher. Model selection and inference: Facts and fiction. *Econometric Theory*, 21(01):21–59, 2005.

- Hannes Leeb and Benedikt M Pötscher. Can one estimate the conditional distribution of post-model-selection estimators? *The Annals of Statistics*, pages 2554–2591, 2006.
- Hannes Leeb and Benedikt M Pötscher. Sparse estimators and the oracle property, or the return of hedges estimator. *Journal of Econometrics*, 142(1):201–211, 2008.
- Vardges Levonyan. What led to the ban on same-sex marriage in california?: Structural estimation of voting data on proposition 8. *Mimeo.*, 2013.
- Zhipeng Liao. Adaptive gmm shrinkage estimation with consistent moment selection. *Econometric Theory*, 29(05):857–904, 2013.
- Richard Lockhart, Jonathan Taylor, Ryan J Tibshirani, and Robert Tibshirani. A significance test for the lasso. *Annals of statistics*, 42(2):413, 2014.
- Kevin Milligan and Marie Rekkas. Campaign spending limits, incumbent spending, and election outcomes. *Canadian Journal of Economics/Revue canadienne d'économique*, 41(4):1351–1374, 2008.
- Sergio Montero. Coalition formation, campaign spending, and election outcomes: Evidence from mexico. *Mimeo*, 2015.
- Hyungsik Roger Moon, Matthew Shum, and Martin Weidner. Estimation of random coefficients logit demand models with interactive fixed effects. Technical report, cemap working paper, Centre for Microdata Methods and Practice, 2014.
- Aviv Nevo. A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy*, 9(4):513–548, 2000.
- Whitney K Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica: Journal of the Econometric Society*, pages 809–837, 1990.
- Whitney K Newey. 16 efficient estimation of models with conditional moment restrictions. *Handbook of statistics*, 11:419–454, 1993.
- Whitney K Newey and Frank Windmeijer. Generalized method of moments with many weak moment conditions. *Econometrica*, 77(3):687–719, 2009.
- Keith T Poole and Howard Rosenthal. A spatial model for legislative roll call analysis. *American Journal of Political Science*, pages 357–384, 1985.
- Marie Rekkas. The impact of campaign spending on votes in multiparty elections. *The Review of Economics and Statistics*, 89(3):573–585, 2007.
- Mathias Reynaert and Frank Verboven. Improving the performance of random coefficients demand models: the role of optimal instruments. *Journal of Econometrics*, 179(1):83–98, 2014.

- Michael L Rothschild. Political advertising: A neglected policy issue in marketing. *Journal of Marketing Research*, pages 58–71, 1978.
- Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- Sara Van de Geer, Peter Bühlmann, Yaacov Ritov, Ruben Dezeure, et al. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.
- Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, pages 894–942, 2010.
- Cun-Hui Zhang and Jian Huang. The sparsity and bias of the lasso selection in high-dimensional linear regression. *The Annals of Statistics*, pages 1567–1594, 2008.
- Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.
- Hui Zou and Runze Li. One-step sparse estimates in nonconcave penalized likelihood models. *Annals of statistics*, 36(4):1509, 2008.

Appendices

A Iterative Computation for Penalty Loadings

The penalized estimators apply data-dependent factor loadings for each of the coefficients included in the model. The data-dependent factor loadings scale the penalty for each coefficient according to the variability of the associated coefficient and the model residual. These loadings appear in Υ_β and Υ_ω in Algorithm 1, Υ_θ in equation (16), Υ_ϕ in equation (18), and Υ_ζ in equation (19). Here we review the application of Belloni et al. (2013a)’s iterative approach to computing these penalty loadings.

A.1 Iterative Computation for Linear Models

Recalling the formula for step I of Algorithm 1:

$$\min_{\beta \in \mathbb{R}^{K_T+1}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_\beta \beta\|_1.$$

As discussed in the details, $\lambda_\beta = 2c\sqrt{JT}\Phi^{-1}(1 - \gamma/2(K_T + 1))$, with Belloni et al. (2013a)’s recommended values being $c = 1.1$ and $\gamma = 0.05/\log(K_T + 1 \vee T)$. The k^{th} diagonal entry in $\hat{\Upsilon}_\beta$ scales the penalty according to the variability in the k^{th} regressor, which we’ll denote $x_{k,jt}$, and the residual $\epsilon_{jt} \equiv \mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p$. The infeasible ideal sets $\hat{\Upsilon}_{\beta,\{k,k\}} = \sqrt{\mathbb{E}[x_{k,jt}^2 \epsilon_{jt}^2]}$. The iterative algorithm A.1 initializes Υ_β with the expected squared value of each regressor, fits the lasso regression, recovers the residuals, and uses these residuals to compute the sample analog to the ideal value. This algorithm extends immediately to Υ_ω . Defining the residual $\varepsilon_{jt} \equiv p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1$, the infeasible ideal penalty values for this problem are $\hat{\Upsilon}_{\omega,\{k,k\}} = \sqrt{\mathbb{E}[x_{k,jt}^2 \varepsilon_{jt}^2]}$. For completeness, the calculation is detailed in Algorithm A.2.

Algorithm A.1 Iterative Algorithm for Υ_β

- I. Initialize $\Upsilon_{k,k}^0 = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1, \dots, K_T$.
 - II. For $\mathcal{I} = 1, \dots, \bar{\mathcal{L}}$, or until $\|\Upsilon^\mathcal{I} - \Upsilon^{\mathcal{I}-1}\| < \delta$:
 - a) Solve $\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^{K_T+1}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\mathcal{S}_{jt} - x'_{0t}\beta_{0j} - x'_{1jt}\beta_1 - p_{jt}\beta_p)^2 + \frac{\lambda_\beta}{T} \|\Upsilon^{\mathcal{I}-1}\beta\|_1$.
 - b) Compute the Residuals: $\hat{\epsilon}_{jt} \equiv \mathcal{S}_{jt} - x'_{0t}\hat{\beta}_{0j} - x'_{1jt}\hat{\beta}_1 - p_{jt}\hat{\beta}_p$.
 - c) Update $\Upsilon_{k,k}^\mathcal{I} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 \hat{\epsilon}_{jt}^2}$, $k = 1, \dots, K_T$.
 - III. Set $\hat{\Upsilon}_\beta = \Upsilon^\mathcal{I}$.
-

Algorithm A.2 Iterative Algorithm for Υ_ω

- I. Initialize $\Upsilon_{k,k}^0 = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1, \dots, K_T$.
 - II. For $\mathcal{I} = 1, \dots, \bar{\mathcal{L}}$, or until $\|\Upsilon^\mathcal{I} - \Upsilon^{\mathcal{I}-1}\| < \delta$:
 - a) Solve $\hat{\omega} = \min_{\omega \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (p_{jt} - x'_{0t}\omega_{0j} - x'_{1jt}\omega_1)^2 + \frac{\lambda_\omega}{T} \|\hat{\Upsilon}_\omega \omega\|_1$.
 - b) Compute the Residuals: $\hat{\epsilon}_{jt} \equiv p_{jt} - x'_{0t}\hat{\omega}_{0j} - x'_{1jt}\hat{\omega}_1$.
 - c) Update $\Upsilon_{k,k}^\mathcal{I} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 \hat{\epsilon}_{jt}^2}$, $k = 1, \dots, K_T$.
 - III. Set $\hat{\Upsilon}_\omega = \Upsilon^\mathcal{I}$.
-

A.2 Iterative Computation for Nonlinear Models

The selection in nonlinear models requires accounting for the additional estimation error introduced by selection on a generated regressor. Consequently, the residual with which to scale the regressor's variability must be augmented by the variance of the generated selection target. Recall the selection problem in equation (18):

$$\tilde{\phi} = \arg \min_{\phi \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J \left(\tilde{\delta}_{jt} - x'_{0t}\phi_{0j} - x'_{1jt}\phi_1 \right)^2 + \frac{\lambda_\phi}{T} \|\hat{\Upsilon}_\phi \phi\|_1.$$

Algorithm A.3 Iterative Algorithm for Υ_ϕ

- I. Initialize $\Upsilon_{k,k}^0 = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1, \dots, K_T$.
 - II. Compute $\hat{\epsilon}_{\delta,jt}^2 = \bar{x}'_{jt} \Sigma_j \bar{x}_{jt} + \sigma_\xi^2$ from the solution to the feasible GMM problem (17).
 - III. For $\mathcal{I} = 1, \dots, \bar{\mathcal{I}}$, or until $\|\Upsilon^\mathcal{I} - \Upsilon^{\mathcal{I}-1}\| < \delta$:
 - a) Solve $\tilde{\phi} = \arg \min_{\phi \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J \left(\tilde{\delta}_{jt} - x'_{0t} \phi_{0j} - x'_{1jt} \phi_{1j} \right)^2 + \frac{\lambda_\phi}{T} \|\Upsilon^{\mathcal{I}-1} \phi\|_1$.
 - b) Compute the Residuals: $\hat{\epsilon}_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t} \hat{\phi}_{0j} - x'_{1jt} \hat{\phi}_{1j}$.
 - c) Update $\Upsilon_{k,k}^\mathcal{I} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 (\hat{\epsilon}_{\phi,jt}^2 + \hat{\epsilon}_{\delta,jt}^2)}$, $k = 1, \dots, K_T$.
 - IV. Set $\hat{\Upsilon}_\phi = \Upsilon^\mathcal{I}$.
-

The Υ_ϕ matrix requires a slight adjustment to account for estimation error in the $\tilde{\delta}_{jt}$'s. Defining

$$\epsilon_{\delta,jt} \equiv \delta_{jt} - \tilde{\delta}_{jt} = \tilde{\delta}_{jt} = \tilde{x}'_{jt} \left(\tilde{\beta}_j - \beta_j \right) + x'_{1jt} \left(\tilde{\beta}_1 - \beta_1 \right) + p_{jt} \left(\tilde{\beta}_p - \beta_p \right) + \tilde{\xi}_{jt} - \xi_{jt}$$

and $\epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x'_{0t} \phi_{0j} - x'_{1jt} \phi_{1j}$, the ideal weight for $\phi_{0j,k}$ is equal to $\sqrt{\bar{E} \left[x_{0t,k}^2 (\epsilon_{\delta,jt} + \epsilon_{\phi,jt})^2 \right]}$ and $\sqrt{\bar{E} \left[x_{1jt,k}^2 (\epsilon_{\delta,jt} + \epsilon_{\phi,jt})^2 \right]}$ for $\phi_{1,k}$.

We can define $\bar{x}_{jt} = [\tilde{x}'_{jt} x'_{1jt}, p_{jt}]'$ and Σ_j as the rows and columns of the variance-covariance matrix for $\tilde{\beta}$ computed using the sandwich covariance matrix from the solution to (17):

$$\hat{\epsilon}_{\delta,jt}^2 = E \left[\epsilon_{\delta,jt}^2 \right] = \bar{x}'_{jt} \Sigma_j \bar{x}_{jt} + \sigma_\xi^2.$$

For feasible implementation, we again initialize the Υ_ϕ matrix with the diagonal variances of the regressors. We then recursively solve (18) to recover the residuals $\epsilon_{\phi,jt}$ and update the Υ_ϕ accordingly.

The approach above doesn't apply as readily to the solution for (19), as we cannot easily characterize the variance of the optimum instruments for the nonlinear features of the

Algorithm A.4 Iterative Algorithm for Υ_ζ

- I. Initialize $\Upsilon_{k,k}^0 = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1, \dots, K_T$.
 - II. For $\mathcal{I} = 1, \dots, \bar{\mathcal{I}}$, or until $\|\Upsilon^\mathcal{I} - \Upsilon^{\mathcal{I}-1}\| < \delta$:
 - a) Solve $\hat{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\tilde{z}_{v,jt} - x'_{0t} \zeta_{0j} - x'_{1jt} \zeta_1)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1$.
 - b) Compute the Residuals: $\hat{\varepsilon}_{\zeta,jt} = \tilde{z}_{v,jt} - x'_{0t} \hat{\zeta}_{0j} - x'_{1jt} \hat{\zeta}_1$.
 - c) Update $\Upsilon_{k,k}^\mathcal{I} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 \hat{\varepsilon}_{\zeta,jt}^2}$, $k = 1, \dots, K_T$.
 - III. Set $\hat{\Upsilon}_\zeta = \Upsilon^\mathcal{I}$.
-

model. However, we don't need to account for the population variance of the asymptotic optimal instruments in our selection of controls. Importantly, the estimated optimal instruments provide the only source of exogenous variation used to identify the heterogeneity in voter impressionability. Consequently, performing selection on the utilized instruments as if they represented the population optimal instruments suffices to control for observable heterogeneity. Recalling the penalization problem:

$$\tilde{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J (\tilde{z}_{v,jt} - x'_{0t} \zeta_{0j} - x'_{1jt} \zeta_1)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1$$

and defining the residual $\varepsilon_\zeta = \tilde{z}_{v,jt} - x'_{0t} \zeta_{0j} - x'_{1jt} \zeta_1$, the ideal $(k, k)^{th}$ entry in $\Upsilon_\zeta = \mathbb{E}[x_{k,j,t}^2 \varepsilon_\zeta^2]$.

We can then apply the approach from Algorithms A.1 and A.2.

A.3 GMM Penalty for Verifying First Order Conditions

While we do not directly evaluate the objective function in the global parameter space for equation (16), we do need to verify the first-order conditions for the local solution based on

the selected model in the last step of Algorithm 2:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q \left(\tilde{\theta}^*, \tilde{x}^k, z, p, s \right) < \lambda_{\theta} v_k, k = 1, \dots, K_0, j = 1, \dots, J.$$

As discussed in the text, the infeasible ideal value of $v_k = \sqrt{\bar{E} [x_{0t,k}^2 \xi_{jt}^2]}$. Here, we are already working from a (putative) local optimum, so we can take the estimated values $\xi \left(\tilde{\theta}^*, \tilde{x}, z, p, s \right)$ to estimate the empirical analog to the expectation:

$$\hat{v}_k = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{J,T} x_{0t,k}^2 \tilde{\xi}_{jt}^2}.$$

This calculation has the added benefit of being computable variable-by-variable to mitigate memory and computational limitations.

B Detailed Statements of Model Assumptions

B.1 Notation

- \bar{E}_T represents the average expectation of a series, for example, $\bar{E}_T [x_{0t} x'_{0t}] = \frac{1}{T} \sum_{t=1}^T \mathbb{E}[x_{0t} x'_{0t}]$.
- \mathcal{E}_T represents the empirical average of a series, for example, $\mathcal{E}_T [x_{0t} x'_{0t}] = \frac{1}{T} \sum_{t=1}^T x_{0t} x'_{0t}$.
- $\phi_{min}(c) \{\Sigma\}$ and $\phi_{max}(c) \{\Sigma\}$ represent the c^{th} smallest and largest eigenvalues of the matrix Σ .

B.2 Assumption 2: Exact Sparsity in Preferences and Spending

As discussed in the main text, assumptions 2.(1)-2.(4) consolidate the restrictions in Belloni et al. (2012) and Belloni et al. (2013a) for the voting application with a fixed number of excluded instruments. The restriction on exponential tails isn't strictly necessary for application in the linear model, as we adopt the Belloni et al. (2012) penalization strategy

Assumption 2 Exact Sparsity in Preferences and Spending

Let $\theta = [\beta'_{01}, \dots, \beta'_{0J}, \beta'_1, \pi'_{01}, \dots, \pi'_{0J}, \pi'_1, \pi'_z]'$. Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^\infty$, has $K_T > T$ possible parameters, $1 \leq k_T < T$ of which are non-zero, where both $K_T \xrightarrow{T \rightarrow \infty} \infty$ and $k_T \xrightarrow{T \rightarrow \infty} \infty$. Further, the number of excluded instruments in z is fixed at $L \geq 2$. Finally, there exists a sequence $\{\delta_T, \Delta_T\} \rightarrow_{T \rightarrow \infty} 0$.

1. The parameter space isn't too large: $\log K_T \leq (T\delta_T)^{1/3}$.
2. The model is sufficiently sparse: $k_T^2 \log^2(K_T \vee T) / T \leq \delta_T$.
 - (a) The number of variables explicitly included, $K_1 + L$ in the model is fixed.
 - (b) In equation (1), the true coefficients have $\sum_{j=1}^J \|\beta_{0j}\|_0 \leq k_T$.
 - (c) In the campaign spending equation (4), the true coefficients have $\sum_{j=1}^J \|\pi_{0j}\|_0 \leq k_T$.
3. Sparse eigenvalues for Gram matrix: There exists a sequence $\ell_T \rightarrow 0$, κ' , and κ'' such that, with probability $1 - \Delta_T$:

$$0 \leq \kappa' \leq \phi_{\min}(\ell_n k_T) \{ \bar{E}_T[x_{0t} x'_{0t}] \} \leq \phi_{\max}(\ell_n k_T) \{ \bar{E}_T[x_{0t} x'_{0t}] \} \leq \kappa'' < \infty.$$

4. Detectable Non-zero Coefficients: $\min\{|\theta_k| \mid \theta_k \neq 0\} > \delta_T$.
5. Exponential tails: There exists $b > 0$ and $r > 0$ such that, for any $\tau > 0$,
 - (a) $P(|\xi_{jt}| > \tau) \leq \exp(-(\tau/b)^r)$;
 - (b) $\forall k, P(|x_{0t,k}| > \tau) \leq \exp(-(\tau/b)^r)$.

that applies moderate deviation theory for self-normalized sums to bound deviations in the maximal element of the score vector. However, we maintain the restriction, as it allows us to use Fan and Liao (2014)'s results, which require results from large deviation theory, in the nonlinear GMM setting.

B.3 Assumption 3: Regularity Conditions for High-Dimensional Linear Logit

The regularity conditions here require somewhat cumbersome notation to specify explicitly. The conditions presented in assumption 3 are sufficient to ensure a law of large numbers and central limit theorem apply to the post-selection estimator with heteroskedastic and non-

Gaussian residuals. With the exponential tails assumption in 2.5, the assumptions restricting the sup-norm of regressors (3.2(a) and 3(b)) can be weakened. In our discussion of the GMM estimator, many of these restrictions are subsumed by simply assuming a uniform law of large numbers applies to the score of the objective function. Assumption 3.3 provide sufficient conditions for such a ULLN to apply in the linear environment. The regularity conditions for the first-stage regression in assumption 3.4 come from Belloni et al. (2012) and ensure the existence of optimal instruments for the endogenous campaign spending and the ability to consistently estimate these instruments via the first-stage regression.

B.4 Assumption 5: Regularity Conditions for GMM Estimator

The GMM regularity conditions are very similar to those in Gillen et al. (2014), reflecting fairly standard restrictions on GMM estimators. The discussion in Caner and Zhang (2013) and Fan and Liao (2014) provide more primitive conditions for these results, though their restrictions are stated for i.i.d. sampling environments, requiring some additional notation to extend to the i.n.i.d. setting here. These assumptions are fairly standard in the literature on GMM estimation, with references to Newey (1990), Newey (1993), Caner (2009), and Newey and Windmeijer (2009).

B.5 Assumption 6: Sparsity Assumptions for High-Dimensional BLP Model

The additional sparsity restrictions for the high-dimensional BLP model are not that different from those in the linear model. The sparsity restriction is a bit tighter, accounting for the need to select variables on estimated optimal instruments. The restricted eigenvalue assumption needs to be extended to the outer product of the gradients, and the exponential tail restriction is extended to the optimal instruments.

Assumption 3 Linear Logit DGP Regularity Conditions

Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^\infty$, has $K_T > T$ possible parameters, $1 \leq k_T < T$ of which are non-zero, where both $K_T \xrightarrow{T \rightarrow \infty} \infty$ and $k_T \xrightarrow{T \rightarrow \infty} \infty$, but the number of excluded instruments in z is fixed at $L \geq 2$. Finally, there exists a sequence $\{\delta_T, \Delta_T\} \rightarrow_{T \rightarrow \infty} 0$ and fixed constants $0 < c < C < \infty$.

1. Sufficient moments for unmodeled variability in the data admit a LLN and CLT:

- (a) $\bar{E} [|\xi_{jt}|^q + |\nu_{jt}|^q] \leq C$,
- (b) $c \leq \bar{E} [\xi_{jt}^2 | x_{jt}, \nu_{jt}] \leq C$, a.s., and,
- (c) $c \leq \bar{E} [\nu_{jt}^2 | x_{jt}, z_{jt}] \leq C$, a.s.

2. Variability in observables and their impact on unobservables is bounded:

- (a) Demographics Controls: $\max_{t \leq T} \|x_{0t}^2\|_\infty k_T T^{(-\frac{1}{2} + \frac{2}{q})} \leq \delta_T$ w.p. $1 - \Delta_T$.
- (b) Candidate Characteristics: $\bar{E} [x_{1jt,k}^q] \leq C$, and $|\beta_{1,k}| < C, k = 1, \dots, K_1$.
- (c) Campaign Expenditure and Impact: $\bar{E} [p_{jt}^q] \leq C$, and $|\beta_p| < C$.

3. Additional regularity restrictions for asymptotic theory with i.n.i.d. sampling:

- (a) $\max_{j \leq J} \bar{E} [|\xi_{jt}|^q] + \bar{E} [|\mathcal{S}_{jt}|^q] + \max_{k \leq K_T} \{\bar{E} [x_{jt,k}^2 \mathcal{S}_{jt}^2] + \bar{E} [x_{jt,k}^3 \xi_{jt}^3] + 1/\bar{E} [x_{jt,k}^2]\} \leq C$.
- (b) $\max_{k \leq K_T, j \leq J} \{ |(\mathcal{E}_T - \bar{E}) [x_{jt,k}^2 \xi_{jt}^2]| + |(\mathcal{E}_T - \bar{E}) [x_{jt,k}^2 \mathcal{S}_{jt}^2]| \} + \max_{t \leq T} \|x_{0t}\|_\infty^2 \frac{k_T \log(T \vee K_T)}{T} \leq \delta_T$.

4. Regularity restrictions for first-stage regression: Let $\tilde{p}_{jt} \equiv p_{jt} - \bar{E} [p_{jt}]$,

- (a) $\max_{j \leq J} \bar{E} [\tilde{p}_{jt}^2] + \max_{l \leq L} \{ \bar{E} [z_{jt,l}^2 \tilde{p}_{jt}^2] + 1/\bar{E} [z_{jt,l}^2 \nu_{jt}^2] \} + \max_{k \leq K_0} \{ \bar{E} [x_{0t,k}^2 \tilde{p}_{jt}^2] + 1/\bar{E} [x_{0t,k}^2 \nu_{jt}^2] \} + \max_{k \leq K_1} \{ \bar{E} [x_{1jt,k}^2 \tilde{p}_{jt}^2] + 1/\bar{E} [x_{1jt,k}^2 \nu_{jt}^2] \} \lesssim 1$.
 - (b) $\max_{j \leq J} \left\{ \max_{k \leq K_0} \bar{E} [x_{0t,k}^3 \nu_{jt}^3] + \max_{k \leq K_1} \bar{E} [x_{1jt,k}^3 \nu_{jt}^3] + \max_{l \leq L} \bar{E} [z_{jt,l}^3 \nu_{jt}^3] \right\} \lesssim \kappa_T$ with $\kappa_T^2 \log^3(K_T \vee T) = o(T)$.
 - (c) $\max_{t \leq T, j \leq J} \left(\max_{l \leq L} z_{jt,l}^2 + \max_{k \leq K_0} x_{0t,k}^2 + \max_{k \leq K_1} x_{1jt,k}^2 \right) [s \log(K_T \vee T)] / T \rightarrow_p 0$.
 - (d) $\max_{j \leq J} \max_{l \leq L} \{ |(\mathcal{E} - \bar{E}) [z_{jt,l}^2 \nu_{jt}^2]| + |(\mathcal{E} - \bar{E}) [z_{jt,l}^2 \tilde{p}_{jt}^2]| \} + \max_{k \leq K_0} \{ |(\mathcal{E} - \bar{E}) [x_{0t,k}^2 \nu_{jt}^2]| + |(\mathcal{E} - \bar{E}) [x_{0t,k}^2 \tilde{p}_{jt}^2]| \} + \max_{k \leq K_1} \{ |(\mathcal{E} - \bar{E}) [x_{1jt,k}^2 \nu_{jt}^2]| + |(\mathcal{E} - \bar{E}) [x_{1jt,k}^2 \tilde{p}_{jt}^2]| \} \rightarrow_p 0$.
-

Assumption 5 Regularity Conditions for GMM Estimator

Let $\theta = [\beta'_{01}, \dots, \beta'_{0J}, \beta'_1, \pi'_{01}, \dots, \pi'_{0J}, \pi'_1, \pi'_z, v_b]'$. For all T , each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^\infty$, satisfies the following restrictions:

1. Compactness of Parameter Set: The true parameter values $\theta_0 \in \Theta_{K_T}$, where $\Theta_K \subset \mathbb{R}^{K_T+2}$ is compact, with a compact limit set $\Theta_\infty \equiv \lim_{T \rightarrow \infty} \Theta_K$.
2. Continuity of Moment Conditions:
 - (a) The unconditional moment condition $E[z_{jt,l} \xi_{jt}(\theta)]$ is continuously differentiable in $\theta, \forall j \leq J, t \leq T$, and, $l \leq L$.
 - (b) For the full-sample moment condition $m_{lT}(\theta) \equiv \bar{E} \left[\frac{1}{J} \sum_{j=1}^J E[z_{jt,l} \xi_{jt}(\theta)] \right]$,
 - i. $m_{lT}(\theta) \xrightarrow{T \rightarrow \infty} m_l(\theta)$ uniformly for $\theta \in \Theta_{K_T}$, for all K_T ,
 - ii. $m_{lT}(\theta)$ is continuously differentiable and its limit $m_l(\theta)$ is continuous in θ , and,
 - iii. $m_{lT}(\theta_0) = 0$ and $m_{lT}(\theta) \neq 0, \forall \theta \neq \theta_0$.
3. Uniform LLN for Sample Analog: Let $g_{jt}(\theta) \equiv [x'_{0t}, x'_{1jt}, z'_{jt}]' \xi_{jt}(\theta)$, the following uniform law of large numbers applies:

$$\sup_{k \leq K_T} \sup_{\theta \in \Theta_k} \left\| \frac{1}{JT} \sum_{j,t=1}^{J,T} (g_{jt}(\theta) - \bar{E}[g_{jt}(\theta)]) \right\| \xrightarrow{T \rightarrow \infty}_p 0.$$

4. Define the $L_T \times K_T$ matrix $\hat{G}_T(\theta) \equiv \frac{1}{JT} \sum_{j,t=1}^{J,T} \frac{\partial g_{jt}(\theta)}{\partial \theta'}$:
 - (a) A uniform law of large numbers holds in a neighborhood of θ_0 for all K_T :

$$\|\hat{G}_T(\theta) - G(\theta)\|_2^2 \xrightarrow{T \rightarrow \infty}_p 0.$$
 - (b) The limiting matrix $G(\theta)$ is continuous in θ and $G(\theta_0)$ has full column rank K_T .
 5. W_T is a positive definite matrix with $\|W_T - W\|_2^2 \rightarrow_p 0$, with W a symmetric, positive definite, and finite matrix.
 6. The expected outer product of the score, $\Omega \equiv \lim_{T \rightarrow \infty} (JT)^{-1} \sum_{j,t=1}^{J,T} E[g_{jt}(\theta) g_{jt}(\theta)']$ is a positive definite, finite matrix.
 7. The minimal and maximal eigenvalues of $\Sigma \equiv G(\theta_0)' \Omega^{-1} G(\theta_0)$, denoted \underline{e} and \bar{e} , are finite and bounded between finite constants $0 < c \leq \underline{e} \leq \bar{e} \leq C < \infty$.
-

Assumption 6 Sparsity Assumptions for High-Dimensional BLP Model

Let $\theta = [\beta'_{01}, \dots, \beta'_{0J}, \beta'_1, \pi'_{01}, \dots, \pi'_{0J}, \pi'_1, \pi'_z v_p]'$. Each data generating process in the sequence $\{\mathcal{P}_T\}_{T=1}^\infty$, has $K_T > T$ possible parameters, $1 \leq k_T < T$ of which are non-zero, where both $K_T \xrightarrow{T \rightarrow \infty} \infty$ and $k_T \xrightarrow{T \rightarrow \infty} \infty$. Further, the number of excluded instruments in z is fixed at

$L \geq 2$. Finally, there exists a sequence $\{\delta_T, \Delta_T\} \rightarrow_{T \rightarrow \infty} 0$ with $\sqrt{\frac{k_T \log k_T}{T}} > \delta_T > \frac{k_T}{T}$.

1. The parameter space isn't too large, with $\log(K_T) = o(T^{-1/3})$.
2. The model is sparse, with the number of non-zero variables, $k_T^3 \log k_T = o(T^{-1})$.
3. The Gram matrix for controls satisfies restricted eigenvalues of Assumption 2.3.
4. The Hessian of the objective function with respect to non-zero variables is almost surely positive definite. There exists a sequence $\ell_T \rightarrow 0$, κ' , and κ'' such that, with probability $1 - \Delta_T$:

$$0 \leq \kappa' \leq \phi_{\min}(\ell_n k_T) \{\Omega\} \leq \phi_{\max}(\ell_n k_T) \{\Omega\} \leq \kappa'' < \infty.$$

5. Non-zero coefficients are bounded away from zero: $\min\{|\theta_k| \text{ s.t. } \theta_k \neq 0\} > 2\delta_T$.
 6. The marginal distributions for controls, instruments, and residual vote shares have exponentially decaying tails.
-