BLP-LASSO for Aggregate Discrete Choice Models with Rich Covariates

Benjamin J. Gillen
Claremont-McKenna College

Sergio Montero
University of Rochester

Hyungsik Roger Moon
University of Southern California
and Yonsei University

Matthew Shum
Caltech

October 9, 2018

Abstract

We introduce the BLP-LASSO model, which augments the classic BLP (Berry, Levinsohn, and Pakes, 1995) random-coefficients logit model to allow for data-driven selection among a high-dimensional set of control variables. 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 preference heterogeneity. 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 that are known to be valid for uniform inferences with respect to variable selection. Our application explores the effect of campaign spending on vote shares in data from Mexican elections.

Keywords. Random-coefficients logit model, High-dimensional regressors, LASSO, Elections, Machine Learning, 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 Gowrisankaran, Chris Hansen, Stefan Holderlein, Ivan Jeliazkov, Dale Poirier, Guillaume Weisang, Frank Windmeijer, and seminar participants at the Advances in Econometrics Conference on Bayesian Model Comparison, the California Econometrics Conference, Emory Political Science Conference, Stanford SITE, ASSA Meetings, UC Irvine, Cal Poly San Luis Obispo, and the University of Arizona.
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 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 address 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 estimation algorithms generalize procedures from Belloni et al. (2012), Belloni et al. (2013), and Chernozhukov et al. (2015) for selecting variables to a nonlinear Berry et al. (1995) model of consumer demand with random coefficients.

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, algorithms of this type are known to identify all the variables necessary for valid inference in the model.

We conduct an empirical investigation of campaign expenditures’ influence on election outcomes, utilizing a structural voting model inspired by discrete-choice demand models from the industrial organization and marketing literatures. We apply our technique to data from Mexican legislative elections. Our analysis yields the robust finding that campaign expenditures significantly influence voter choices.

The model we propose, which we call “BLP-LASSO”, is, in essence, the BLP random-coefficients logit model (Berry et al., 1995) augmented for variable selection. We review the BLP model in
section 4 and, as a benchmark, present results using the Mexican data with a set of pre-specified demographic variables for controls.

Section 5 introduces and describes our BLP-LASSO model. Estimation and inference here pose some conceptual and computational challenges. We provide two double-LASSO type algorithms that are adapted to the BLP demand model estimation. The BLP-LASSO results for the effects of campaign expenditures, the main endogenous variable of interest, are similar to the estimates from the pre-selected model. However, the demographic covariates selected by the BLP-LASSO procedure are quite distinct, and in some cases surprising, compared to our pre-selected variables. This points toward the value of integrated data-driven variable selection in the estimation of BLP-style discrete-choice demand models.

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 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 represent 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, 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 incorporates some form of $L_1$-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 the true 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 $L_0$ or $L_1$ 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 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 naively-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) 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. (2013) 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.

\(^1\)Generalized 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. (2013) characterizing inference under an approximate sparsity condition that constrains the error in a sparse representation of the true data generating process.
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. Chernozhukov et al. (2015) and its appendix consider general GMM problems. 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.

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 Rosen-
thal (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 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. 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.” Our application is closely related to Montero (2016)’s structural analysis of the incentives for coalition formation in Mexican elections, and utilizes the same data as that study.

Beyond structural approaches, a massive body of empirical research investigates the influence of campaign expenditures on vote shares using natural and field experiments. These studies 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 in these environments.

3 Voting in Mexico

Our empirical application explores the effect of campaign spending on voting in Mexican legislative elections. Mexico is a federal republic with the executive branch headed by the president and legislative power wielded by a bicameral Congress. We restrict attention to elections for the lower chamber of Congress, known as the Chamber of Deputies, which are held every three years. While voting is mandatory for all citizens aged 18 and older, there are no sanctions in place enforcing participation. Legislators can be re-elected only in non-consecutive terms, limiting the potential for incumbency advantage in these elections.

The Chamber of Deputies has 500 total members. Seats in the chamber are contested under
a mixed electoral system. The country is divided by the national electoral authority (INE) into 300 electoral districts (see Figure 1)—seeking to equalize representation while preserving state boundaries—and in each district candidates compete in a first-past-the-post race for a single seat in the chamber (i.e., the winner is elected by a simple plurality of votes). For these district races, candidates can be nominated by a single political party or by a coalition of multiple parties. Election laws also allow candidates to run independently or as a write-in campaign, although their vote shares are negligible. The remaining 200 seats in the chamber are assigned to registered political parties according to a proportional representation (PR) rule. Specifically, votes are pooled by party across all districts and each party receives a share of the 200 PR seats proportional to the number of votes received by that party’s candidates.\(^2\) To identify candidates for the PR assignment, parties submit national lists of up to 200 candidates concurrent with registering district candidates. However, parties must secure at least 2\% of the national vote to retain their national accreditation. Importantly, the mixed nature of the electoral system mitigates incentives to vote strategically: a vote for an uncompetitive candidate in a district race is not a wasted vote as it impacts the nominating party’s performance in the PR component of the election.

![Figure 1: Mexican states (shaded) and electoral districts (delimited)](image)

---

\(^2\)Additional 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.
3.1 Fundraising and Advertising in Mexican Elections

Campaign funding for Mexican parties is allocated from the federal budget. Of this total allocation, 30% is divided equally among all registered 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 campaign expenditures. Consequently, candidate fundraising is negligible in these elections, with the party national committees supplying the financial and administrative resources to run individual 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. Media advertising is highly regulated 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 public funding, with 30% divided equally and the remaining 70% proportionally to the parties’ national vote shares in the most recent Chamber of Deputies election. While airtime is free, parties expend campaign resources to produce their own TV and radio ads.

3.2 Parties and Coalitions in the 2012 Chamber of Deputies Election

Our analysis focuses on the 2012 Chamber of Deputies election. Seven political parties participated in the election. Two parties, the National Action Party (PAN) and the New Alliance Party (NA), participated independently, nominating individual candidates in every district. Three parties, the Party of the Democratic Revolution (PRD), the Labor Party (PT), and the Citizens’ Movement (MC), formed a total coalition called the Progressive Movement (MP), nominating a common candidate in each district. And the two remaining parties, the Institutional Revolutionary Party (PRI) and the Ecologist Green Party of Mexico (PVEM), formed a partial coalition called Commitment for Mexico (CM), joining forces in only 199 districts. As such, the 2012 Chamber of Deputies election featured five competing candidates in those districts where PRI and PVEM candidates ran independently, and four candidates where one ran as part of the CM coalition.

The 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.
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 are shown in parentheses.

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

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, we treat coalitions as a single party, but we do control for the actual party affiliation of coalition candidates as a characteristic. Notably, party leaders must register coalition agreements before the electoral authority at least 6 months prior to the election. These agreements specify the party affiliation of each candidate to be put forth in every district. Individual candidates, however, are selected and formally nominated only 3-4 months ahead of the election, following primary elections or other appointment mechanisms by local district committees that are beyond the direct control of national party leaders.

3.3 Electoral and Census Data

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. Voter turnout rates are also available by district.

To control for observable heterogeneity in voter preferences, we have access to rich demographic and socioeconomic data—over 200 variables—from the 2010 population census, which the National Statistics and Geography Institute (INEGI) makes available at the electoral-district level. This includes district breakdowns by age and gender, education, religion, language, marital status, disability, healthcare services, employment, public infrastructure and services, as well as various
proxies for household income or wealth.

Campaign spending data are self-reported by the parties to the electoral authority, subject 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 1 reports summary district-level spending statistics by party or coalition, broken down by the coalition structure of the district, which determines the number of competing candidates. To characterize the geographic dispersion of campaign spending, Figure 3 maps the distribution of 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. In particular, the variability in expenditures between parties is greater than would be expected by differences in, e.g., the price of media, which would affect all parties equally.

<table>
<thead>
<tr>
<th>Party</th>
<th>Districts with distinct PRI, PVEM candidates</th>
<th></th>
<th>Districts with joint CM candidate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. dev.</td>
<td>Mean</td>
</tr>
<tr>
<td>PRI</td>
<td>54.9</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>PVEM</td>
<td>18.3</td>
<td>7.6</td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td></td>
<td></td>
<td>83.5</td>
</tr>
<tr>
<td>PAN</td>
<td>38.0</td>
<td>10.4</td>
<td>42.1</td>
</tr>
<tr>
<td>MP</td>
<td>56.4</td>
<td>19.7</td>
<td>55.4</td>
</tr>
<tr>
<td>NA</td>
<td>19.7</td>
<td>8.5</td>
<td>17.2</td>
</tr>
</tbody>
</table>
4 A random-utility model of voting

In the literature on demand estimation, a common approach to capturing heterogeneity in preferences allows for random coefficients in the individual’s utility model, leading to the Berry et al. (1995) (“BLP”) random-coefficients logit model. We incorporate this here by allowing voters to have heterogeneous impressionability, introducing random coefficients 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 electorate.

We begin by reviewing the BLP model without addressing the variable selection step necessary to address the high dimensionality of the problem. In district $t$, we observe vote shares based on individual voters (indexed by $i$) who choose from among the candidates (indexed by $j = 1, \ldots, J$) competing in the district. We represent the option to not vote or to write in a non-party candi-
date as an “outside good” indexed by \( j = 0 \). To characterize the preferences of a representative voter, 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 \( p_{jt} \). Finally, we allow unobserved candidate characteristics or quality to affect voters’ preferences through a product-market-specific latent shock, \( \xi_{jt} \).

To model heterogeneity in preferences, let an individual voter’s preference for candidate \( j \) be defined as:

\[
    u_{ijt} = x_{0t} \beta_{0j} + x_{1jt}' \beta_1 + p_{jt} \beta_p + p_{jt} b_{ip} + \xi_{jt} + \epsilon_{ijt}, \quad b_{ip} \sim N \left( 0, v_p^2 \right).
\]

(1)

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

\[
    Pr \{ y_{ijt} = j | b_i \} = \frac{\exp \left\{ x_{0t} \beta_{0j} + x_{1jt}' \beta_1 + p_{jt} \beta_p + p_{jt} b_{ip} + \xi_{jt} \right\}}{1 + \sum_{r=1}^J \exp \left\{ x_{0t} \beta_{0r} + x_{1rt}' \beta_1 + p_{rt} \beta_p + p_{rt} b_{ip} + \xi_{rt} \right\}}.
\]

(2)

We integrate equation 2 to compute the expected vote share for a candidate in the district. Letting \( \Phi \) denote the standard normal distribution’s cumulative density, candidate \( j \)'s expected vote share in district \( t \) can be written as:

\[
    s_{jt} = \int Pr \{ y_{ijt} = j | b_i \} \, d\Phi \left( \frac{b_i}{v_p} \right).
\]

(3)

Following BLP, we “invert” the expected vote share equation (3) to recover the party-district specific shocks. Given any value of the parameters, \( \theta \), and observed vote shares, \( s \), 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 \( \theta \), instruments \( z_{jt} \) are orthogonal to these shocks, i.e., \( E[\xi_{jt} (\theta, X, p, s) | z_{jt}] = 0 \), so that \( \theta \) is estimated by minimizing a GMM objective function with weighting matrix \( W \):

\[
    Q (\theta, x, z, p, s) = \frac{1}{J^T} \xi (\theta, X, p, s)' z W z' \xi (\theta, X, p, s),
\]

(4)

For 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.
**Assumption 1 BLP 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 (1).

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 (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 (8).

4. Residual shocks to expected voter preferences in a district are endogenously correlated with unmodeled variation in campaign spending, as in equation (9), but has a zero conditional expectation given district and candidate characteristics and observable instruments: 
   \[ E[\xi_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0. \]

where $\xi(\theta, X, p, s)$ is the vector consisting of $\xi_{jt}(\theta, X, p, s)$ and $z$ is the matrix that stacks $z_{jt}$. In the standard setting with a fixed number of controls and instruments, for any positive-definite $W$, minimizing equation (4) provides an asymptotically normal estimator for the parameters in $\theta$. 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 (4). 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 2 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 1 and 2 are sufficient to establish the usual consistency and asymptotic normality results for the value of $\theta$ that minimizes equation (4).

### 4.1 Results: BLP model with Fixed Controls

We begin our empirical analysis of heterogeneous impressionability of voters in Mexico using a pre-specified set of controls considered in Table 2. Table 3’s Panel A reports the expected coefficients
Assumption 2 Regularity Conditions for GMM Estimator

1. Compactness of parameter space: The true parameter values $\theta_0 \in \Theta_{K_T}$, where $\Theta_{K_T} \subset \mathbb{R}^{K_T+2}$ is compact, with a compact limit set $\Theta_\infty \equiv \lim_{T \to \infty} \Theta_{K_T}$.

2. Continuity and differentiability of sample-analog and population moment conditions in parameter space.

3. Letting $g_{jt}(\theta) \equiv [x_{0t}^t, x_{1jt}^t, z_{jt}^t]^T \xi_{jt}(\theta)$, a uniform law of large numbers ensures the sample-analog, $\frac{1}{JT} \sum_{j,t=1}^{J,T} g_{jt}(\theta)$, converges to the population moment condition.

4. A uniform law of large numbers applies to the Hessian of the sample analog to the population moment condition, $\hat{G}_T(\theta) \equiv \frac{1}{JT} \sum_{j,t=1}^{J,T} \frac{\partial g_{jt}(\theta)}{\partial \theta}$.

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, $\Omega \equiv \lim_{T \to \infty} (JT)^{-1} \sum_{j,t=1}^{J,T} \mathbb{E}[g_{jt}(\theta) g_{jt}(\theta)^T]$, is a positive definite, finite matrix.

7. The matrix $\Sigma \equiv G(\theta_0)^T \Omega^{-1} G(\theta_0)$ is almost surely positive definite and finite.

Table 2: Pre-Selected Controls for Fixed Model of Voting

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

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 is associated with a party-specific fixed effect, $x_{0t}$, in the utility model.

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. Panel B reports the significance of the demographic controls included in the model with heterogeneous impressionability. We see that only a small fraction of these interactions are significant at usual significance levels, suggesting that these interactions may not be the most relevant ones for capturing heterogeneity in vote shares across districts. Next, we turn to a data-driven procedure to select more relevant controls.
Table 3: Results: BLP Model with Pre-selected Controls

<table>
<thead>
<tr>
<th>Panel A: Main Results</th>
<th>Coefficient</th>
<th>Std Error</th>
<th>t-Stat</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditures</td>
<td>0.77</td>
<td>0.38</td>
<td>2.04</td>
<td>4%</td>
</tr>
<tr>
<td>Expenditures²</td>
<td>-0.04</td>
<td>0.03</td>
<td>-1.39</td>
<td>17%</td>
</tr>
<tr>
<td>Panel B: Significance of Demographic Controls (p-Values)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Party</td>
<td>Region 1</td>
<td>Region 2</td>
<td>Region 3</td>
<td>Region 4</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>-----------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>MP</td>
<td>15%</td>
<td>0%</td>
<td>84%</td>
<td>35%</td>
</tr>
<tr>
<td>NA</td>
<td>2% *</td>
<td>11%</td>
<td>52%</td>
<td>41%</td>
</tr>
<tr>
<td>PAN</td>
<td>6% *</td>
<td>0%</td>
<td>4%</td>
<td>10%</td>
</tr>
<tr>
<td>PRI</td>
<td>62%</td>
<td>46%</td>
<td>89%</td>
<td>28%</td>
</tr>
<tr>
<td>PVEM</td>
<td>11%</td>
<td>43%</td>
<td>1%</td>
<td>95%</td>
</tr>
<tr>
<td>CXM</td>
<td>32%</td>
<td>98%</td>
<td>10%</td>
<td>29%</td>
</tr>
</tbody>
</table>

Panel A reports the return to campaign expenditures and squared campaign expenditures using the nonlinear BLP voting model with heterogeneous impressionability estimated from equation (4) with interactive fixed-effects between the political party and demographic controls listed in Table 2. 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.
5 BLP-LASSO: Variable Selection in the BLP Voting Model

We now address the implications of the high-dimensional setting for 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 $\theta$ for which the residual shock to preferences, $\xi_{jt}$, can be equal to zero for all observations. 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.

Inference requires solving a penalized GMM objective function with a LASSO penalty. We apply a data-dependent penalization that is robust to heteroskedasticity in sampling across markets:

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{K_T+2}} Q (\theta, x, z, p, s) + \frac{\lambda_\theta}{T} \| \Upsilon_\theta \|_1. \quad (5)$$

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

5.1 Implementing Variable Selection via Penalized GMM

Since it is computationally infeasible to directly optimize the GMM objective function in extremely high-dimensional problems, we now describe the approach we use for selecting variables using a penalized GMM estimator. Our procedure for estimating this model proceeds by running two algorithms in sequence, which we call Algorithm 1 and Algorithm 2 and summarize in the tables below.
5.1.1 First Step: Multinomial Logit Without Random Coefficients

Algorithm 1 essentially applies the double-LASSO procedure in Belloni et al. (2012) and Belloni et al. (2013) to a (multinomial) logit voting model without random coefficients. The utility specification here is

\[ u_{ijt} = x_{0t}^t \beta_0 + x_{1jt}^t \beta_1 + p_{jt} \beta_p + \xi_{jt} + \epsilon_{ijt}, \]  

leading to a log-linear expression for aggregate vote shares:

\[ S_{jt} \equiv \log s_{jt} - \log s_{0t} = x_{0t}^t \beta_0 + x_{1jt}^t \beta_1 + p_{jt} \beta_p + \xi_{jt}. \]  

We supplement this with a linear equation for campaign expenditures

\[ p_{jt} = x_{0t}^t \pi_0 + x_{1jt}^t \pi_1 + z_{jt}^t \pi_z + \nu_{jt}, \quad \mathbb{E}[\nu_{jt}|x_{0t}, x_{1jt}, z_{jt}] = 0, \]  

and a variance components structure for the utility shock \( \xi_{jt} \):

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

Since the structural equations for this model take the form of linear regressions (equations (7) and (8)), the procedure in the Belloni et al. papers can be directly applied. This consists of two stages of penalized estimation for selecting control variables followed by a two-stage-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 coefficients. This is summarized in Algorithm 1.

5.1.2 Second Step: Multinomial Logit With Random Coefficients

Algorithm 2 takes as initial input the demographic controls \( \tilde{x} \) chosen by Algorithm 1. Using these controls, we first optimize the GMM objective function for the BLP model without any penalization:

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

16
the asymptotic covariance matrix of \( \tilde{\lambda} \) requires a slight adjustment to account for estimation error in the \( \tilde{\lambda} \).

Given the solution \( \tilde{\beta} \), we can recover the latent mean utilities:

\[
\tilde{\delta}_{jt} = \bar{x}_{jt}' \tilde{\beta} + x_{1jt}' \tilde{\beta}_1 + p_{jt} \tilde{\beta}_p + \xi_{jt}.
\]

These provide the outcome variable for which we need to select the relevant demographic controls using another application of the LASSO. Let \( x^{III} \equiv \{ x | \tilde{\phi}(x) \neq 0 \} \), where:

\[
\tilde{\phi} = \arg \min_{\phi \in \mathbb{R}^{K_T}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\delta_{jt} - x_{otj} \phi_{0j} - x_{1jt}' \phi_1)^2 + \lambda_\phi \| \tilde{\Upsilon}_\phi \phi \|_1.
\] (11)

The penalization term \( \lambda_\phi \) has the same expected form as previous applications. The \( \Upsilon_\phi \) 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 \( \epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x_{otj} \phi_{0j} - x_{1jt}' \phi_1 \), the ideal weight for \( \zeta_{0j,k} \) is equal to \( \sqrt{E \left[ x_{0l,k}^2 (\epsilon_{\delta,jt} + \epsilon_{\phi,jt})^2 \right]} \), and \( \sqrt{E \left[ x_{1lj,k}^2 (\epsilon_{\delta,jt} + \epsilon_{\phi,jt})^2 \right]} \) for \( \beta_{1,k} \). The additional residuals can be characterized by using the asymptotic covariance matrix of \( \tilde{\theta} \), which can be consistently estimated using the sandwich
covariance matrix from the penalized GMM estimator.

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 \( \delta_{jt} = \tilde{x}_{0t} \hat{\beta}_j + \tilde{x}_{1jt} \hat{\beta}_1 + p_{jt} \hat{\beta}_p + \xi_{jt} (\tilde{\theta}, x, z, p, s) \), where \( \tilde{\theta} \equiv [\hat{\beta}_1, \ldots, \hat{\beta}_j, \hat{\beta}_1, \hat{\delta}_p]^T \):

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

III. Estimate the generalized endogenous regressor of the heterogeneity in impressionability. Compute the derivative of the moment condition with respect to the variability parameter \( v_p \):

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

IV. Select controls for mean utilities. Let \( x^{III} = x^I \cup x^II \) and compute the unpenalized GMM estimate:

\[
\tilde{x}^* = \arg \min_\phi \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\tilde{\delta}_{jt} - \tilde{x}_{0t} \phi_{0j}^T - \tilde{x}_{1jt} \phi_{1})^2 + \lambda_\phi \|\hat{\Phi}\|_1.
\]

V. Select controls for \( \tilde{z}_{v,jt} \). Let \( x^{IV} = x^{III} \cup x^{IV} \) and compute the unpenalized GMM estimate:

\[
\tilde{x}^* = \arg \min_\zeta \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\tilde{z}_{v,jt} - \tilde{x}_{0t} \zeta_{0j} - \tilde{x}_{1jt} \zeta_{1})^2 + \lambda_\zeta \|\hat{\zeta}\|_1.
\]

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

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

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 that the first order improvement in the objective function from including this variable \( x_{0k} \) for any party is dominated by the penalty:

\[
q_k = \frac{\partial}{\partial \phi_{0j}} Q(\tilde{\theta}^*, \tilde{x}^k, z, p, s) < \lambda_\phi \Upsilon_{(k,\tilde{\theta}^*,(k)_1)} k = 1, \ldots, K, j = 1, \ldots, J.
\]

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 \( x^V = \{x_k : q_k > \lambda_\phi \Upsilon_{(k,\tilde{\theta}^*,(k)_1)} \} \). 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: \( \lambda_\phi = \lambda_\zeta = 2c \sqrt{T} \Phi^{-1} (1 - \gamma/(2K)) \) and \( \lambda_\theta = 2c \sqrt{T} \Phi^{-1} (1 - \gamma/(2K + 8)) \), with \( c = 1.1 \) and \( \gamma = \frac{0.05}{\text{log}(K + T)} \). The details for calculating the diagonal factor loading matrices \( \hat{\Upsilon}_{(k)} \), 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.

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 (10), we compute the derivative of the objective function with respect to the variance of the random coefficients, $v_p$:

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

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 constraints for the MPEC objective function. We can interpret $\tilde{z}_{v,jt}$ as an approximated endogenous regressor that corresponds to the heterogeneity of the preference. The next step is to select demographic control variables that explain the approximated endogenous regressor.

Let $x^{IV} \equiv \{ x \mid \tilde{\zeta}(x) \neq 0 \}$, where:

$$\tilde{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\tilde{z}_{v,jt} - x_{0t}\zeta_{0j} - x_{1jt}\zeta_{1j})^2 + \frac{\lambda_{\zeta}}{T} \parallel \hat{\Upsilon}_{\zeta} \zeta \parallel_1.$$

(12)

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 (12), as in $\Upsilon_{\phi}$. 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 $\Upsilon_{\zeta}$, we ignore the generated-regressors problem.

5.2 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

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

(13)

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:

\[ \tilde{z}_{p,t} = \tilde{x}_f' \hat{\pi}_x + z_f' \hat{\pi}_z. \]

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:

\[ \theta = \arg \min_{\theta} Q (\theta, \tilde{x}^*, \tilde{z}, p, s). \]  \hfill (14)

The last step of Algorithm 2 then verifies that this solution also satisfies the first-order conditions for the penalized objective function (5) 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 constraints in the MPEC objective function:

\[ q_k = \partial_{\beta_{0jk}} Q (\tilde{\theta}^*, \tilde{x}^k, z, p, s) < \lambda_{\theta v_k, j} = 1, \ldots, J, \]

where \( \tilde{x}^k = \tilde{x}^* \cup x_{0k} \) for each excluded demographic control \( x_{0k} \). 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.

It is known that the post-selection estimator defined in Algorithm 1 is asymptotically normal uniformly, which can be used in conducting uniformly valid inference (e.g., see Belloni et al. (2012) and Chernozhukov et al. (2015)). The post-selection estimator of the BLP coefficients in Algorithm 2 is an extension of the post-selection method for linear IV models. In this paper, we mainly focus on the empirical and computational aspects of this problem, and we do not provide an asymptotic theory that justifies the procedure in Algorithm 2. We leave it as a future research topic.
<table>
<thead>
<tr>
<th>Expected Coefficients</th>
<th>Coefficient</th>
<th>Std Error</th>
<th>t-Stat</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditures</td>
<td>0.465</td>
<td>0.198</td>
<td>2.35</td>
<td>2%</td>
</tr>
<tr>
<td>Expenditures²</td>
<td>-0.020</td>
<td>0.015</td>
<td>-1.33</td>
<td>18%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance of Coefficients</th>
<th>Coefficient</th>
<th>Std Error</th>
<th>t-Stat</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expenditures</td>
<td>0.05</td>
<td>0.67</td>
<td>0.07</td>
<td>94%</td>
</tr>
<tr>
<td>Expenditures²</td>
<td>0.00</td>
<td>0.06</td>
<td>0.01</td>
<td>99%</td>
</tr>
</tbody>
</table>

### Panel B: Selected Demographic Controls

<table>
<thead>
<tr>
<th>MP</th>
<th>NA</th>
<th>PAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Party FE (+)**</td>
<td>Party FE (-)</td>
<td>Party FE (-)**</td>
</tr>
<tr>
<td>Female Popn &gt;60 (-)**</td>
<td>% Female HoH (-)</td>
<td>% Female HoH (-)**</td>
</tr>
<tr>
<td>% Pop 18-24 (-)</td>
<td>% Pop Married (+)</td>
<td>% Pop Married (+)</td>
</tr>
<tr>
<td>% Pop w/Elementary Ed (-)*</td>
<td>% Pop w/Elementary Ed (-)*</td>
<td>% Pop w/Elementary Ed (-)*</td>
</tr>
</tbody>
</table>

### Panel A: Main Results

5.3 Results: BLP-LASSO model

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 $\gamma \log (K_T \lor T) = 0.10$.

Panel A of Table 4 reports the impact of campaign expenditures on mean utilities. The estimated first-order mean effect of 0.465 is a bit lower than the pre-selected model’s result of 0.767. 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 4 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 ideology 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.

6 Conclusion

We present several results on high-dimensional inference and apply these techniques to an empirical analysis of voting behavior in Mexican elections. Our analysis involves estimating aggregate demand models with a very large number of demographic covariates. Though our statistical analysis is largely informed by previously established properties of high-dimensional inference techniques, their extensions to our 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. While 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 (2016) 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.
References


Algorithm A.1 Iterative Algorithm for $\hat{\Upsilon}_\beta$

I. Initialize $\hat{\Upsilon}_\beta^0 = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1, \ldots, K_T$.

II. For $I = 1, \ldots, \bar{I}$, or until $\|\hat{\Upsilon}_I - \hat{\Upsilon}_{I-1}\| < \delta$:

   a) Solve $\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^{K_T+1}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (S_{jt} - x_{0t}^j \beta_0 - x_{1jt}^j \beta_1 - p_{jt}^j \beta_p)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_{I-1} \beta\|_1$.

   b) Compute the Residuals: $\hat{\epsilon}_{jt} \equiv S_{jt} - x_{0t}^j \hat{\beta}_0 - x_{1jt}^j \hat{\beta}_1 - p_{jt}^j \hat{\beta}_p$.

   c) Update $\hat{\Upsilon}_{k,k} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 \hat{\epsilon}_{jt}^2}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\beta = \hat{\Upsilon}_I$.

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 $\hat{\Upsilon}_\beta$ and $\hat{\Upsilon}_\omega$ in Algorithm 1, $\hat{\Upsilon}_\theta$ in equation (5), $\hat{\Upsilon}_\phi$ in equation (11), and $\hat{\Upsilon}_\zeta$ in equation (12). Here, we review the application of Belloni et al. (2013)’s iterative approach to computing these penalty loadings.

A.1 Iterative Computation for Linear Models

Recalling the formula for step I of Algorithm 1:

$$
\min_{\beta \in \mathbb{R}^{K_T+1}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (S_{jt} - x_{0t}^j \beta_0 - x_{1jt}^j \beta_1 - p_{jt}^j \beta_p)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_{I-1} \beta\|_1.
$$

As discussed in the details, $\lambda_\beta = 2c\sqrt{JT} \Phi^{-1} \left(1 - \gamma/2(K_T + 1)\right)$, with Belloni et al. (2013)’s recommended values being $c = 1.1$ and $\gamma = 0.05\log(K_T + 1)$. 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 $x_{k,j,t}$, and the residual $\epsilon_{jt} \equiv S_{jt} - x_{0t}^j \hat{\beta}_0 - x_{1jt}^j \hat{\beta}_1 - p_{jt}^j \hat{\beta}_p$.

The infeasible ideal sets $\hat{\Upsilon}_{\beta,(k,k)} = \sqrt{\mathbb{E}[x_{k,j,t}^2 \epsilon_{jt}^2]}$. The iterative algorithm A.1 initializes $\hat{\Upsilon}_\beta$ 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 $\hat{\Upsilon}_\omega$. Defining the residual $\epsilon_{jt} \equiv p_{jt} - x_{0t}^j \omega_0 - x_{1jt}^j \omega_1$, the infeasible ideal penalty values for this problem are $\hat{\Upsilon}_{\omega,(k,k)} = \sqrt{\mathbb{E}[x_{k,j,t}^2 \epsilon_{jt}^2]}$. For completeness, the calculation is detailed in Algorithm A.2.

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
Algorithm A.2 Iterative Algorithm for $\Upsilon_\omega$

I. Initialize $\Upsilon_{0,k} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{J,T} x_{k,j,t}^2}$, $k = 1, \ldots, K_T$.

II. For $I = 1, \ldots, \tilde{I}$, or until $\|\Upsilon_I - \Upsilon_{I-1}\| < \delta$:

a) Solve $\tilde{\omega} = \min_{\omega \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (p_{jt} - x_{0t}^j \omega_{0j} - x_{1jt}^j \omega_1)^2 + \frac{\lambda_\omega}{T} \|\hat{\Upsilon}_\omega \tilde{\omega}\|_1$.

b) Compute the Residuals: $\tilde{\varepsilon}_{jt} = p_{jt} - x_{0t}^j \tilde{\omega}_{0j} - x_{1jt}^j \tilde{\omega}_1$.

c) Update $\Upsilon_{I,k} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{J,T} x_{k,j,t}^2 \tilde{\varepsilon}_{jt}^2}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\omega = \Upsilon_I$.

must be augmented by the variance of the generated selection target. Recall the selection problem in equation (11):

$$\hat{\phi} = \arg \min_{\phi \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\tilde{\delta}_{jt} - x_{0t}^j \phi_{0j} - x_{1jt}^j \phi_1)^2 + \frac{\lambda_\phi}{T} \|\hat{\Upsilon}_\phi \phi\|_1.$$  

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

$$\epsilon_{\delta,jt} = \delta_{jt} - \hat{\delta}_{jt} = \tilde{x}_{jt}^j \left( \tilde{\beta}_j - \tilde{\beta}_j \right) + x_{1jt}^j \left( \tilde{\beta}_1 - \tilde{\beta}_1 \right) + p_{jt} \left( \tilde{\beta}_p - \tilde{\beta}_p \right) + \tilde{\xi}_{jt} - \xi_{jt}$$

and $\epsilon_{\phi,jt} = \tilde{\delta}_{jt} - x_{0t}^j \phi_{0j} - x_{1jt}^j \phi_1$, the ideal weight for $\phi_{0j,k}$ is equal to $\sqrt{\mathbb{E} \left[ x_{0t,k}^2 \left( \epsilon_{\delta,jt} + \epsilon_{\phi,jt} \right)^2 \right]}$ and

$$\sqrt{\mathbb{E} \left[ x_{1jt,k}^2 \left( \epsilon_{\delta,jt} + \epsilon_{\phi,jt} \right)^2 \right]} \text{ for } \phi_{1,k}.$$

We can define $\tilde{x}_{jt} = [\tilde{x}_{jt}^j x_{1jt}^j p_{jt}]'$ and $\Sigma_j$ as the rows and columns of the variance-covariance matrix for $\hat{\beta}$ computed using the sandwich covariance matrix from the solution to (10):

$$\tilde{\varepsilon}_{\delta,jt}^2 = \mathbb{E} \left[ \epsilon_{\delta,jt}^2 \right] = \tilde{x}_{jt}^j \Sigma_j \tilde{x}_{jt} + \sigma_\xi^2.$$  

For feasible implementation, we again initialize the $\Upsilon_\phi$ matrix with the diagonal variances of the regressors. We then recursively solve (11) to recover the residuals $\epsilon_{\phi,jt}$ and update the $\Upsilon_\phi$ accordingly.

The approach above doesn’t apply as readily to the solution for (12), as we cannot easily characterize the variance of the optimum instruments for the nonlinear features of the 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:

$$\tilde{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\tilde{\varepsilon}_{v,jt} - x_{0t}^j \zeta_{0j} - x_{1jt}^j \zeta_1)^2 + \frac{\lambda_\zeta}{T} \|\hat{\Upsilon}_\zeta \zeta\|_1$$

and defining the residual $\varepsilon_{\zeta} = \tilde{\varepsilon}_{v,jt} - x_{0t}^j \zeta_{0j} - x_{1jt}^j \zeta_1$, the ideal $(k,k)^{th}$ entry in $\Upsilon_{\zeta} = \mathbb{E} \left[ x_{k,jt}^2 \varepsilon_{\zeta}^2 \right]$. We can then
Algorithm A.3 Iterative Algorithm for $\Upsilon_\phi$

I. Initialize $\Upsilon_0^k,k = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1,\ldots,K_T$.

II. Compute $\hat{\epsilon}_j^2 = \hat{x}_j^2 \Sigma_j \hat{x}_j + \sigma^2_\xi$ from the solution to the feasible GMM problem (10).

III. For $I = 1,\ldots,\bar{I}$, or until $\|\Upsilon_I - \Upsilon_{I-1}\| < \delta$:
   a) Solve $\hat{\phi} = \arg \min_{\phi \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^T \sum_{j=1}^J \left( \hat{\delta}_{jt} - x_{0t} \phi_0 - x_{1jt} \phi_1 \right)^2 + \lambda \phi \|\Upsilon_{I-1}\|_1$.
   b) Compute the Residuals: $\hat{\epsilon}_{\phi,jt} \equiv \hat{\delta}_{jt} - x_{0t} \hat{\phi}_0 - x_{1jt} \hat{\phi}_1$.
   c) Update $\Upsilon_{I}^{k,k} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 \left( \hat{\epsilon}_{\phi,jt}^2 + \hat{\epsilon}_{\delta,jt}^2 \right)}$, $k = 1,\ldots,K_T$.

IV. Set $\hat{\Upsilon}_\phi = \Upsilon_I$.

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 (5), 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:

$$q_k \equiv \frac{\partial}{\partial \theta_{0,jk}} Q(\hat{\theta}, \tilde{x}_0^k, z, p, s) < \lambda_\theta v_k, k = 1,\ldots,K_0, j = 1,\ldots,J.$$

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

$$\hat{v}_k = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{0t,k}^2 \xi_{jt}^2}.$$

This calculation has the added benefit of being computable variable-by-variable to mitigate memory and computational limitations.
Algorithm A.4 Iterative Algorithm for $\Upsilon_\zeta$

I. Initialize $\Upsilon_0^{k,k} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2}$, $k = 1, \ldots, K_T$.

II. For $I = 1, \ldots, \bar{I}$, or until $\|\Upsilon_I - \Upsilon_{I-1}\| < \delta$:
   
   a) Solve $\hat{\zeta} = \arg \min_{\zeta \in \mathbb{R}^{KT}} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} (\tilde{z}_{v,jt} - x'_{0t}\hat{\zeta}_{0j} - x'_{1jt}\hat{\zeta}_{1j})^2 + \frac{\lambda}{T} \|\hat{\Upsilon}_I \zeta\|_1$.

   b) Compute the Residuals: $\hat{\epsilon}_{\zeta,jt} = \tilde{z}_{v,jt} - x'_{0t}\hat{\zeta}_{0j} - x'_{1jt}\hat{\zeta}_{1j}$.

   c) Update $\Upsilon_I^{k,k} = \sqrt{\frac{1}{JT} \sum_{j,t=1}^{JT} x_{k,j,t}^2 \hat{\epsilon}_{\zeta,jt}^2}$, $k = 1, \ldots, K_T$.

III. Set $\hat{\Upsilon}_\zeta = \Upsilon_I$.