The Econometrics Journal

Econometrics Journal (2019), volume **22**, pp. 262–281.
doi: 10.1093/ectj/utz010

BLP-2LASSO for aggregate discrete choice models with rich covariates

BENJAMIN J. GILLEN[†], SERGIO MONTERO[‡], HYUNGSIK ROGER MOON[§] AND MATTHEW SHUM[‖]

[†]*Claremont-McKenna College, Robert Day School of Economics and Finance, Claremont, CA 91711.*
Email: Benjamin.Gillen@claremontmckenna.edu

[‡]*University of Rochester, Dept. of Political Science, Rochester, NY 14627.*
Email: smontero@rochester.edu

[§]*University of Southern California, Dept. of Economics, Los Angeles, CA 90089; Yonsei University, Dept. of Economics, Seoul, Republic of Korea.*
Email: moonr@usc.edu

[‖]*California Institute of Technology, Division of Humanities and Social Sciences, Pasadena, CA 91125.*
Email: mshum@caltech.edu

First version received: 11 October 2018; final version accepted: 14 January 2019.

Summary We introduce the BLP-2LASSO 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 using the 'double-LASSO' procedure proposed by Belloni, Chernozhukov, and Hansen (2013). Economists often study consumers' aggregate behaviour 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 from the recent literature in high-dimensional econometrics. 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*.

JEL codes: *C55*.

1. INTRODUCTION

When analysing 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 in firms' pricing policies. The abundance of such variables, whether from census data, localized search trends, or local media viewership

© 2019 Royal Economic Society. Published by Oxford University Press. All rights reserved. For permissions please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

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 analysing 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 incorporate the so-called 'double-LASSO' procedures from Belloni et al. (2012) and Belloni et al. (2013) (see also a recent survey by Belloni et al. 2017).[1]

The specific problem of interest addresses high-dimensional demographic data for local markets that may help to characterize local preferences. To address this problem, we adopt techniques proposed in 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 the influence of campaign expenditures 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-2LASSO', is, in essence, a combination of the Berry et al. (1995) ('BLP') random-coefficients logit model and the 'double-LASSO' variable selection procedures developed in Belloni et al. (2012) and Belloni et al. (2013). 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-2LASSO 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-2LASSO 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-2LASSO procedure are quite distinct, and in some cases surprising, compared with our pre-selected variables. This points towards 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

[1] Relatedly, Chernozhukov et al. (2015) study linear equation models with many instruments and controls, including an application to a stylized BLP (Berry et al. 1995) demand model.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

organization and marketing. Finally, the statistical techniques we apply utilize recent innovations in automated machine learning developing 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. The LASSO estimator of Tibshirani (1996) 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 through the SCAD penalty of Fan and Li (2001), the elastic net of Zou and Hastie (2005), the bridge estimator of Huang et al. (2008), the infeasible LASSO of Bickel et al. (2009), and the minimax concave penalty of Zhang (2010). This literature has also inspired several closely related estimators, including the Dantzig selector of Candes and Tao (2007) and the feasible Dantzig selector of Gautier and Tsybakov (2011), as well as the square-root LASSO of Belloni et al. (2011). Each of these estimators incorporates some form of L_1-regularization to the objective function's maximization problem, which leads to a solution with a large number of zero coefficients.

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.[2] If an estimator classifies zero and non-zero coefficients with perfect accuracy as the sample grows, the estimator satifies an oracle property.

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) 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 owing 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 that the relevant controls are included in the econometric model before performing inference in an unpenalized post-selection model.

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 analysing 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), Fan and Liao (2014), Chernozhukov et al. (2015) and its appendix, and Luo and Chernozhukov (2016). Several of these papers, including Chernozhukov et al. (2015) and its appendix and Luo and Chernozhukov (2016), address the issue of moment selection, as in Andrews (1999) and Andrews and Lu (2001).

[2] 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.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Readers can refer to Belloni et al. (2017) for an extensive survey on the estimation and inference of high-dimensional models along with applications to various econometrics problems.

Our BLP-2LASSO model builds directly on the analysis of Gillen et al. (2014) of demand models with complex products. That paper provides a large set of simulation results which highlight how well the BLP-2LASSO model performs relative to other approaches.

2.2. *Structural Models of Campaign Spending and Voting*

The empirical analysis of voting data presents a particularly challenging exercise for political scientists owing to the large number of factors driving voter behaviour, the endogeneity induced by party competition and candidate selection, and behavioural phenomena driving individual voter decisions. Starting 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 because of the 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 analysing political data. Discrete-choice approaches to analysing 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 that voter's behaviour. The problem we consider is closest to that of 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 procedure for addressing the inherent ambiguity in determining which of these estimates is best supported by the data. Our application is closely related to the structural analysis of Montero (2016) 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 influence voter behaviour. 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.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

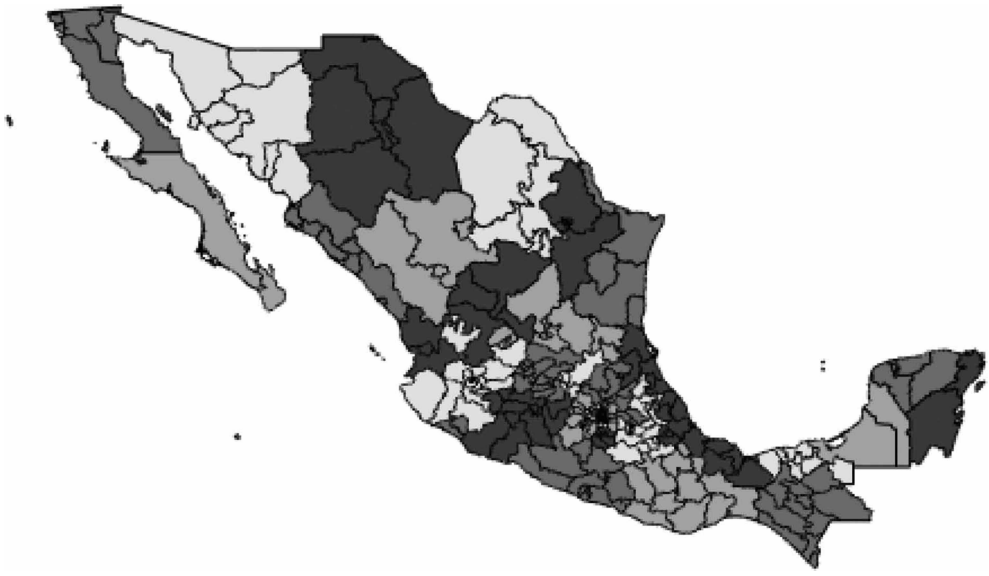

Figure 1. Mexican states (shaded) and electoral districts (delimited).

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 a total of 500 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.[3] 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.

[3] 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.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

3.1. Fundraising and Advertising in Mexican Elections

Campaign funding for Mexican parties is allocated from the federal budget.[4] 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 that 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 up to 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 advertisements.

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.

Figure 2 presents the parties on a one-dimensional ideology spectrum based on a national poll by Consulta Mitofsky (2012) in 2012, along with their national vote shares in the 2012 election. In analysing 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 six 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 three to four 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.

[4] 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 totalled about U.S.$250 million in 2012.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

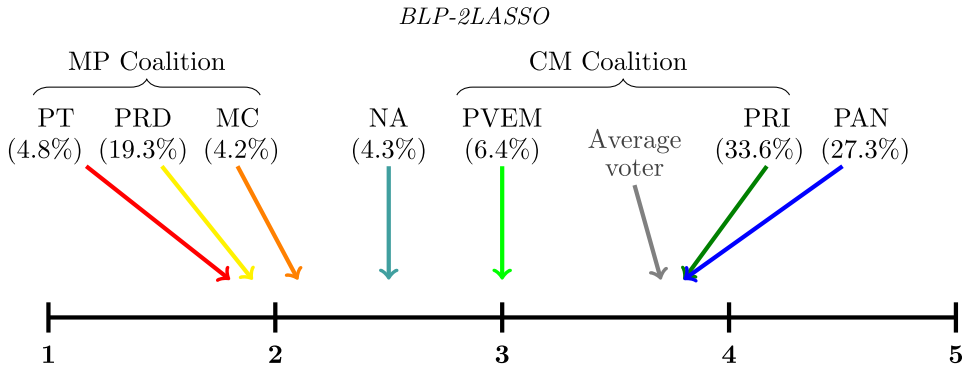

Figure 2. Left–right ideological identification of Mexican parties and voters.
Source: Consulta Mitofsky (2012). A total of 1,000 registered voters were asked in December 2012 to place the parties and themselves on a five-point, left–right ideology scale. Arrows point to national averages. Parties' vote shares are shown in parentheses.

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 by 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 *over*reported 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 neighbouring 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, for example, the price of media, which would affect all parties equally.

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

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

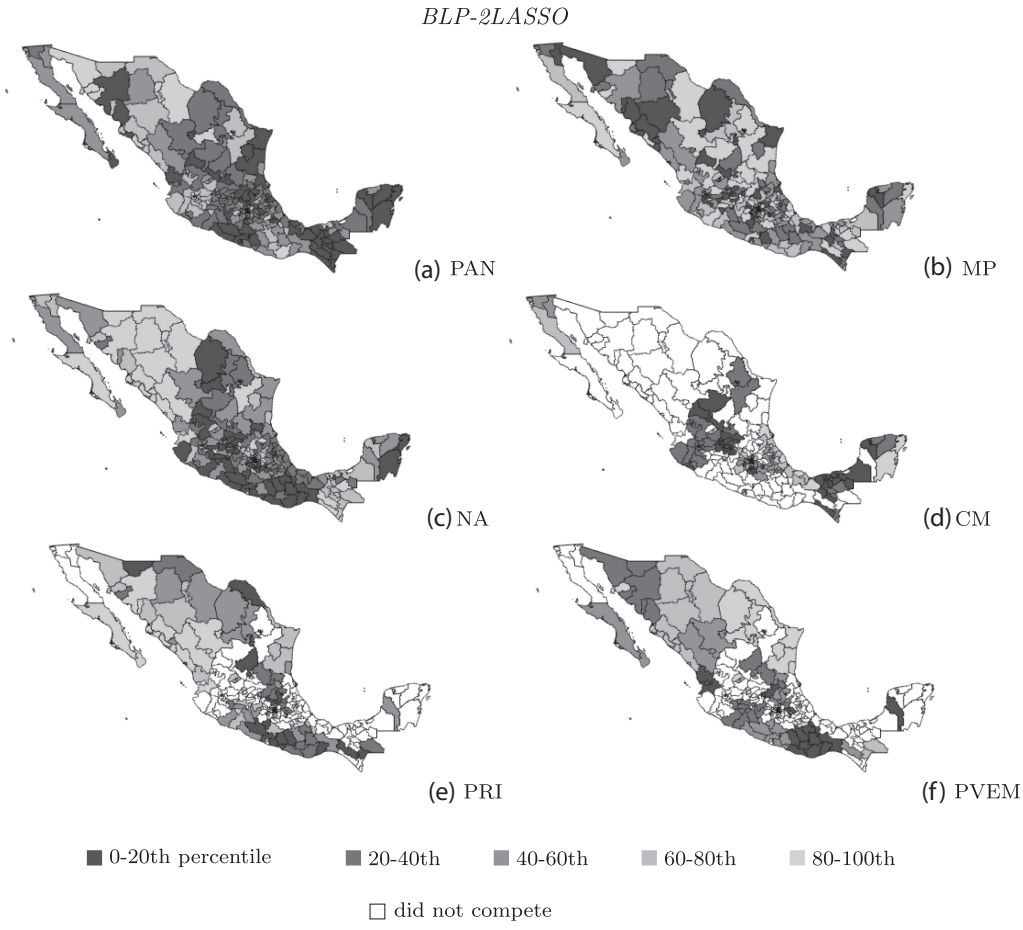

Figure 3. Geographic distribution of campaign spending by party.

Table 1. Campaign spending (in thousands of U.S. dollars).

	Districts with distinct PRI, PVEM candidates		Districts with joint CM candidate	
Party	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

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

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 candidate 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}.[5] Finally, we allow unobserved candidate characteristics or quality to affect voters' preferences through a product-market-specific latent shock, ξ_{jt}.

To model heterogeneity in preferences, let an individual voter's preference for candidate j be defined as

$$u_{ijt} = x_{0t}'\beta_{0j} + x_{1jt}'\beta_1 + p_{jt}\beta_p + p_{jt}b_{ip} + \xi_{jt} + \epsilon_{ijt}, \quad b_{ip} \sim N\left(0, v_p^2\right). \tag{4.1}$$

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:

$$Pr\left\{y_{ijt} = j|b_{ip}\right\} = \frac{\exp\left\{x_{0t}'\beta_{0j} + x_{1jt}'\beta_1 + p_{jt}\beta_p + p_{jt}b_{ip} + \xi_{jt}\right\}}{1 + \sum_{r=1}^{J}\exp\left\{x_{0t}'\beta_{0r} + x_{1rt}'\beta_1 + p_{rt}\beta_p + p_{rt}b_{ip} + \xi_{rt}\right\}}. \tag{4.2}$$

We integrate Equation (4.2) to compute the expected vote share for a candidate in the district. Letting Φ denote the standard normal distribution's CDF, candidate j's expected vote share in district t can be written as

$$s_{jt} = \int Pr\left\{y_{ijt} = j|b_{ip}\right\} d\Phi\left(b_{ip}/v_p\right). \tag{4.3}$$

Following BLP, we 'invert' the expected vote share (Equation 4.3) to recover the party-district-specific shocks. Given any value of the parameters, $\theta = (\beta_{01}', \ldots, \beta_{0J}', \beta_1', \beta_p, v_p)'$, and observed vote shares, Berry et al. (1995) show that a contraction mapping recovers these shocks, which we denote $\xi_{jt}(\theta, X, p, s)$, where X, p and s are the matrix and vectors that stack (x_{0t}', x_{1jt}'), p_{jt} and s_{jt}. Under the true values for θ, instruments z_{jt} are orthogonal to these shocks, i.e., $\mathbb{E}\left[\xi_{jt}\left(\theta, X, p, s\right)|z_{jt}\right] = 0$, so that θ is estimated by minimizing a GMM objective function with weighting matrix W:

$$Q\left(\theta, x, z, p, s\right) = \frac{1}{JT}\xi\left(\theta, X, p, s\right)' zWz'\xi\left(\theta, X, p, s\right), \tag{4.4}$$

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 (4.4) provides an asymptotically normal estimator for the parameters in θ. These asymptotic properties are discussed in Berry et al. (2004). To address numerical issues in the evaluation of this estimator, Dube et al. (2012) present an MPEC (Mathematical Programming with Equilibrium Constraints) algorithm, which we also use here.

[5] 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 a quadratic influence on voter latent utilities.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Table 2. Pre-selected controls for fixed model of voting.

Regional dummies	Demographics	Economic status
Region 1	Percentage of pop. age 18–24	Unemployment
Region 2	Percentage of pop. age 65+	Percentage of households w/car
Region 3	Percentage of pop. that is married	Percentage of households w/refrigerators
Region 4	Average years of education	Percentage of households w/o basic utils
Region 5	Percentage of pop. with elementary ed.	Percentage 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 is associated with a party-specific fixed effect, x_{0t}, in the utility model.

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 (4.4). Gandhi and Houde (2015) demonstrate that vote shares themselves can serve as valuable instruments. Reynaert and Verboven (2014) illustrate how sensitive the estimator is to implementation with Chamberlain (1987) optimal instruments, particularly with respect to estimating the variance parameters v_p. Our implementation adopts the Chamberlain (1987) approach, as presented in the appendix of Nevo (2000), since these instruments are easily recovered from the gradient of the constraints in the MPEC algorithm.

4.1. Results: BLP model with Fixed Controls

We begin our empirical analysis of the heterogeneous impressionability of voters in Mexico using a pre-specified set of controls considered in Table 2. Panel A of Table 3 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. 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.

5. BLP-2LASSO: 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 nontrivial 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. 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

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Table 3. Results: BLP model with pre-selected controls.

Panel A: Main results

Expected coefficients		Coefficient	Std error	t-stat	p-value
Expenditures		0.79	0.39	2.06	4%
Expenditures^2		−0.04	0.03	−1.36	17%
Variance of coefficients		Coefficient	Std error	t-stat	p-value
Expenditures		0.12	0.41	0.30	77%
Expenditures^2		0.00	0.04	0.00	100%

Panel B: Significance of demographic controls (p-values*100)

Party	Party FE	Region 1	Region 2	Region 3	Region 4
MP	0**	47	0**	0**	20
NA	0**	96	0**	43	4*
PAN	0**	46	31	5	90
PRI	0**	54	99	87	28
PVEM	0**	6	10	18	14
CXM	0**	34	29	30	63

Party	Unempl.	Car	Percentage of households with Refrigerator	Utilities	Female head
MP	58	28	0**	44	83
NA	26	0**	19	14	10
PAN	23	64	0**	0**	75
PRI	41	3*	23	11	6
PVEM	31	9	52	75	40
CXM	33	93	61	99	5

Party	Pop. 18–24	Pop. 65+	Married	Element ed. +	Avg years school
MP	74	1*	14	3*	47
NA	68	83	77	10	2*
PAN	5*	2*	2*	11	8
PRI	36	98	35	14	78
PVEM	2*	30	62	61	4*
CXM	55	41	14	72	58

Panel A reports the return to campaign expenditures and squared campaign expenditures using the nonlinear BLP voting model with heterogeneous impressionability estimated from Equation (4.4) 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.

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:

$$\tilde{\theta} = \arg\min_{\theta} Q(\theta, x, z, p, s) + \frac{\lambda_\theta}{T} \|\hat{\Upsilon}_\theta \theta\|_1. \tag{5.5}$$

The properties of such a regularized GMM method are studied in Belloni et al. (2017) (cf. Example 7 there).

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

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 5.1 and Algorithm 5.2 and summarize in the tables below.

5.1.1. First Step: Double-LASSO to obtain initial set of controls. In Algorithm 5.1, we select an initial set of controls for the BLP-2LASSO model by ignoring the random coefficients. Without random coefficients, the utility specification is

$$u_{ijt} = x_{0t}'\beta_{0j} + x_{1jt}'\beta_1 + p_{jt}\beta_p + \xi_{jt} + \epsilon_{ijt}, \tag{5.6}$$

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

$$\mathcal{S}_{jt} \equiv \log s_{jt} - \log s_{0t} = x_{0t}'\beta_{0j} + x_{1jt}'\beta_1 + p_{jt}\beta_p + \xi_{jt}. \tag{5.7}$$

We supplement this with a linear equation for campaign expenditures:

$$p_{jt} = x_{0t}'\pi_{0j} + x_{1jt}'\pi_1 + z_{jt}'\pi_z + \nu_{jt}, \quad \mathbb{E}\left[\nu_{jt}|x_{0t}, x_{1jt}, z_{jt}\right] = 0, \tag{5.8}$$

and a variance components structure for the utility shock ξ_{jt}:

$$\xi_{jt} = \rho\nu_{jt} + \eta_{jt}, \quad \mathbb{E}\left[\eta_{jt}|\nu_{jt}\right] = 0. \tag{5.9}$$

Since the structural equations for this model take the form of linear regressions (Equations 5.7 and 5.8), the procedure in the Belloni et al. papers can be directly applied. This consists of two stages of penalized estimation for selecting control variables. The two stages of selection reflect our need to model conditional expectations for the expected impact of control variables on both (I) the vote-share outcome and (II) the campaign-spending treatment variable. 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 5.1.

5.1.2. Second Step: Estimating the BLP-2LASSO model. Algorithm 5.2 takes as initial input the demographic controls $\tilde{x}$ chosen by Algorithm 5.1. Using these controls, we first optimize the GMM objective function for the BLP model without any penalization:

$$\tilde{\theta} = \arg\min Q\left(\theta, \tilde{x}, z, p, s\right). \tag{5.10}$$

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

$$\tilde{\delta}_{jt} = \tilde{x}_{0t}'\tilde{\beta}_j + \tilde{x}_{1jt}'\tilde{\beta}_1 + p_{jt}\tilde{\beta}_p + \tilde{\xi}_{jt}.$$

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

$$\tilde{\phi} = \underset{\phi\in\mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x_{0t}'\phi_{0j} - x_{1jt}'\phi_1\right)^2 + \frac{\lambda_\phi}{T}\|\hat{\Upsilon}_\phi\phi\|_1. \tag{5.11}$$

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 $\epsilon_{\delta,jt} \equiv \delta_{jt} - \tilde{\delta}_{jt}$

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Algorithm 5.1 Selecting initial set of controls for BLP-2LASSO model.

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

$$\tilde{\beta}_I = \underset{\beta \in \mathbb{R}^{K_T+1}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\mathcal{S}_{jt} - x_{0t}'\beta_{0j} - x_{1jt}'\beta_1 - p_{jt}\beta_p\right)^2 + \frac{\lambda_\beta}{T} \|\hat{\Upsilon}_\beta \beta\|_1. \tag{5.9}$$

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

$$\tilde{\omega} = \underset{\omega \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(p_{jt} - x_{0t}'\omega_{0j} - x_{1jt}'\omega_1\right)^2 + \frac{\lambda_\omega}{T} \|\hat{\Upsilon}_\omega \omega\|_1. \tag{5.9}$$

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 $\sqrt{\bar{E}\left[x_{k,jt}^2 \epsilon_{jt}^2\right]}$, 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.

and $\epsilon_{\phi,jt} \equiv \tilde{\delta}_{jt} - x_{0t}'\phi_{0j} - x_{1jt}'\phi_1$, the ideal weight for $\zeta_{0j,k}$ is equal to $\sqrt{\bar{E}\left[x_{0t,k}^2 \left(\epsilon_{\delta,jt} + \epsilon_{\phi,jt}\right)^2\right]}$, and $\sqrt{\bar{E}\left[x_{1jt,k}^2 \left(\epsilon_{\delta,jt} + \epsilon_{\phi,jt}\right)^2\right]}$ for $\beta_{1,k}$. The additional residuals are characterized by the asymptotic covariance matrix of $\tilde{\theta}$, which can be consistently estimated using the sandwich covariance matrix from the penalized GMM estimator.

Given the relatively minor adjustment to the penalty term, the selected controls in x^{III} typically match the input demographic controls $\tilde{x}$. We further note that, by applying Algorithm 5.1, 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 heterogeneity of impressionability, selecting on features that were not included in Algorithm 5.1. 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 (5.10), we compute the derivative of the objective function with respect to the variance of the random coefficients, v_p:

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

The formula for $\tilde{z}_{v,jt}$ from Berry et al. (1999) is presented in the appendix in Nevo (2000) 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 preferences.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Algorithm 5.2 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 $\tilde{x} = x^I \bigcup x^{II}$ as the controls for a homogeneous model.**

II. **Compute GMM estimates for heterogeneous model using selected controls.** Let $\tilde{\delta}_{jt} \equiv \tilde{x}_{0t}'\tilde{\beta}_{0j} + \tilde{x}_{1jt}'\tilde{\beta}_1 + p_{jt}\tilde{\beta}_p + \xi_{jt}\left(\tilde{\theta}, x, z, p, s\right)$, where $\tilde{\theta} \equiv (\tilde{\beta}_{01}', \ldots, \tilde{\beta}_{0J}', \tilde{\beta}_1', \tilde{\beta}_p, \tilde{v}_p)'$:

$$\tilde{\theta} = \arg\min_{\theta} Q\left(\theta, \tilde{x}, z, p, s\right). \tag{5.11}$$

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:

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

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

$$\tilde{\phi} = \arg\min_{\phi \in \mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{\delta}_{jt} - x_{0t}'\phi_{0j} - x_{1jt}'\phi_1\right)^2 + \frac{\lambda_\phi}{T}\|\hat{\Upsilon}_\phi\phi\|_1. \tag{5.11}$$

V. **Select controls for $\tilde{z}_{v,jt}$.** Let $x^{IV} \equiv \left\{x | \tilde{\zeta}(x) \neq 0\right\}$, where:

$$\tilde{\zeta} = \arg\min_{\zeta \in \mathbb{R}^{K_T}} \frac{1}{JT}\sum_{t=1}^{T}\sum_{j=1}^{J}\left(\tilde{z}_{v,jt} - x_{0t}'\zeta_{0j} - x_{1jt}'\zeta_1\right)^2 + \frac{\lambda_\zeta}{T}\|\hat{\Upsilon}_\zeta\zeta\|_1. \tag{5.11}$$

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

$$\tilde{\theta}^* = \arg\min_{\theta} Q\left(\theta, \tilde{x}^*, z, p, s\right). \tag{5.11}$$

VII. **Verify first-order conditions in unselected model.** For each excluded demographic control x_{0k}, define $\tilde{x}^k = \tilde{x}^* \bigcup x_{0k}$. Verify that the first-order improvement in the objective function from including this variable x_{0k} for any party is dominated by the penalty:

$$q_k \equiv \frac{\partial}{\partial \beta_{0jk}} Q\left(\tilde{\theta}^*, \tilde{x}^k, z, p, s\right) < \lambda_\theta \Upsilon_{\theta,(k,k)}, k = 1, \ldots, K_0, j = 1, \ldots, J. \tag{5.11}$$

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 $x^V = \left\{x_k : q_k > \lambda_\theta \Upsilon_{\theta,(k,k)}\right\}$. Redefine $\tilde{x} = \tilde{x}^* \bigcup x^V$ and return to Step (II) until there are no changes in the set of included variables.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

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 $\gamma = \frac{0.05}{\log(K_T \vee T)}$. The details for calculating the diagonal factor loading matrices $\hat{\Upsilon}_{(\cdot)}$, 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.

The next step then selects demographic control variables that explain the approximated endogenous regressor. Let $x^{IV} \equiv \left\{x | \tilde{\zeta}(x) \neq 0\right\}$, where

$$\tilde{\zeta} = \underset{\zeta \in \mathbb{R}^{K_T}}{\arg\min} \frac{1}{JT} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\tilde{z}_{v,jt} - x_{0t}'\zeta_{0j} - x_{1jt}'\zeta_1\right)^2 + \frac{\lambda_\zeta}{T} \| \hat{\Upsilon}_\zeta \zeta \|_1. \tag{5.12}$$

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 (5.12), 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.

5.2. *Post-Selection Inference via Unpenalized GMM*

Combining the selected controls from Algorithm 5.1, $\tilde{x}$, with x^{III} and x^{IV}, define $\tilde{x}^* = \tilde{x} \bigcup x^{III} \bigcup x^{IV}$. We then compute the unpenalized, post-selection estimator

$$\tilde{\theta}^* = \underset{\theta}{\arg\min}\, Q\left(\theta, \tilde{x}^*, z, p, s\right). \tag{5.13}$$

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:

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

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

$$\theta = \underset{\theta}{\arg\min}\, Q\left(\theta, \tilde{x}^*, \tilde{z}, p, s\right). \tag{5.14}$$

The last step of Algorithm 5.2 then verifies that this solution also satisfies the first-order conditions for the penalized objective function (5.5) 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

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

gradients can be recovered from the Jacobian of the constraints in the MPEC objective function:

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

where $\tilde{x}^k = \tilde{x}^* \bigcup x_{0k}$ 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 instrumental variables 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.

5.3. *Results: BLP-2LASSO model*

Even with only validating local optimality conditions for the selected model, practical implementation of Algorithm 5.2 is quite computationally intensive. In analysing the Mexican voting data, fitting a single penalty specification requires approximately 80 core-hours of computation. Owing to the intensive computational resources required for estimation, we focus our empirical analysis on the benchmark penalty specification where $c = 1.1$ and $\gamma \log\left(K_T \vee T\right) = 0.10$.

Panel A of Table 4 reports the impact of campaign expenditures on mean utilities. The estimated first-order mean effect is 0.753, which is quite close to the pre-selected model's result of 0.79. The standard error of the model with variable selection is slightly lower than that of the pre-specified model, 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. Interestingly, these demographic controls were most important for characterizing voter preferences for the largest independent party (PAN) and parties that are part of a coalition. 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 behaviour 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.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Table 4. Results: BLP-2LASSO model.

Panel A: Main results				
Expected coefficients	Coefficient	Std error	t-stat	p-value*100
Expenditures	0.753	0.325	2.32	2
Expenditures^2	−0.043	0.028	−1.53	12
Variance of coefficients	Coefficient	Std error	t-stat	p-value*100
Expenditures	0.155	0.256	0.61	55
Expenditures^2	0.000	0.029	0.00	99
Panel B: Selected demographic controls				
NA	**MP**	**PVEM**		
Party FE (–)**	Party FE (–)**	Party FE (–)**		
Pop w/ISTE benefits (–)	% Region 2 (–)**	Sq occupants/room in dwelling (–)		
	% Region 4 (+)**			
	% pop density (+)			
PAN	% Sq pop w/Car (–)** (–)*	**CXM**		
Party FE (–)**	Sq pop non-Christian religion (+)**	Party FE (–)**		
% Region 2 (+)**	Sq % Live w/parents (–)**	Region 2 dummy (–)		
Pop Catholic (+)		Region 4 dummy (+)		
Sq % live w/parents (+)**	**PRI**	Pop w/ISTE benefits (+)		
Pop in one-room dwellings (+)	Party FE (–)**	Sq Pop w/ISTE benefits (+)		
% HH with internet (–)	Pop density (–)**	Indigenous pop (+)**		

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.

Our assumption of a linear campaign financing rule may not be consistent with all models 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 does not affect variable selection and inference. Another strategy might adopt a control function approach to estimation, perhaps following the strategy in Kawai (2014) of incorporating techniques from production function estimation.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

We owe special thanks to Alexander Charles Smith for important insights early in developing the project. We are grateful for comments from David Brownstone, Martin Burda, Garland Durham, Jeremy Fox, Gautam Gowrisankaran, 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, Emory Political Science Conference, Stanford SITE, ASSA and APSA meetings, UC Irvine, Cal Poly San Luis Obispo, and the University of Arizona.

REFERENCES

Andrews, D. W. (1999). Consistent moment selection procedures for generalized method of moments estimation. *Econometrica 67*, 543–63.

Andrews, D. W. and B. Lu (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics 101*, 123–64.

Belloni, A., D. Chen, V. Chernozhukov and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*, 2369–429.

Belloni, A., V. Chernozhukov, D. Chetverikov, C. Hansen and K. Kato (2017). High-dimensional econometrics and regularized GMM. *Handbook of Econometrics*, to appear.

Belloni, A., V. Chernozhukov and C. Hansen (2014). Inference on treatment effects after selection amongst high-dimensional controls. *Review of Economic Studies*, *81*, 608–50.

Belloni, A., V. Chernozhukov and L. Wang (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming, *Biometrika 98*, 791–806.

Berry, S., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*, 841–90.

Berry, S., J. Levinsohn and A. Pakes (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review 89*(3), 400–30.

Berry, S., O. B. Linton and A. Pakes (2004). Limit theorems for estimating the parameters of differentiated product demand systems. *The Review of Economic Studies 71*, 613–54.

Bickel, P. J., Y. Ritov and A. B. Tsybakov (2009). Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, *37*, 1705–32.

Bradic, J., J. Fan and W. Wang (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*, 325–49.

Candes, E. and T. Tao (2007). The Dantzig selector: statistical estimation when p is much larger than n. *The Annals of Statistics*, *35*, 2313–51.

Caner, M. (2009). Lasso-type GMM estimator. *Econometric Theory 25*, 270–90.

Caner, M. and H. H. Zhang (2013). Adaptive elastic net for generalized methods of moments. *Journal of Business and Economic Statistics*, *32*, 30–47.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics 34*, 305–34.

Che, H., G. Iyer and R. Shanmugam (2007). Negative advertising and voter choice. Technical report, working paper, University of Southern California, Los Angeles, CA.

Cheng, X. and Z. Liao (2015). Select the valid and relevant moments: An information-based lasso for GMM with many moments. *Journal of Econometrics 186*, 443–64.

Chernozhukov, V., C. Hansen and M. Spindler (2015). Post-selection and post-regularization inference in linear models with many controls and instruments. *American Economic Review 105*(5), 486–90.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Consulta Mitofsky(2012). Geometría electoral en México (electoral geometry in Mexico). Public Opinion Poll, Retrieved from http://www.consulta.mx/web/images/MexicoOpina/2013/NA_GEOMETRIA_ELECTORAL.pdf (in Spanish).

Dube, J.-P., J. Fox and C.-L. Su (2012). Improving the numerical performance of BLP static and dynamic discrete choice random coefficients demand estimation. *Econometrica*, *80*, 2231–69.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*, 1348–60.

Fan, J. and Y. Liao (2014). Endogeneity in high dimensions. *Annals of Statistics 42*, 872–917.

Fan, J. and J. Lv (2011). Nonconcave penalized likelihood with np-dimensionality. *IEEE Transactions on Information Theory 57*, 5467–84.

Gandhi, A. and J. Houde (2015). Measuring substitution patterns in differentiated products industries? The missing instruments, Working paper.

Gautier, E. and A. Tsybakov (2011). High-dimensional instrumental variables regression and confidence sets. (arXiv:1105.2454).

Gerber, A. (1998). Estimating the effect of campaign spending on senate election outcomes using instrumental variables. *American Political Science Review 92*, 401–11.

Gerber, A. (2011). Field experiments in political science. In *Handbook of Experimental Political Science*, edited by J. Druckman, D. Green, J. Kuklinski and A. Lupia , Cambridge UNiversity Press (Cambridge, UK), 115–40.

Gillen, B. J., H. R. Moon and M. Shum (2014). Demand estimation with high-dimensional product characteristics. *Advances in Econometrics: Bayesian Model Comparison*34, 301–24.

Gordon, B. R. and W. R. Hartmann (2013). Advertising effects in presidential elections. *Marketing Science 32*, 19–35.

Gordon *et al*(2012). Marketing and politics: Models, behavior, and policy implications. *Marketing Letters 23*, 391–403.

Green, D. P. and J. S. Krasno (1988). Salvation for the spendthrift incumbent: Reestimating the effects of campaign spending in house elections. *American Journal of Political Science*, *32* 884–907.

Horowitz, J. L. and J. Huang (2010). The adaptive lasso under a generalized sparsity condition. *Working paper, Northwestern University*.

Huang, J., J. L. Horowitz and S. Ma (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. *The Annals of Statistics 36*, 587–613.

Jacobson, G. C. (1978). The effects of campaign spending in congressional elections. *American Political Science Review 72*, 469–91.

Jacobson, G. C. (1985). Money and votes reconsidered: Congressional elections, 1972–1982. *Public Choice 47*, 7–62.

Kawai, K. (2014). Campaign finance in US house elections, Working paper.

King, G. (1997). *A solution to the ecological inference problem*, Harvard University Press.

Leeb, H. and B. M. Pötscher (2005). Model selection and inference: Facts and fiction. *Econometric Theory 21*, 21–59.

Leeb, H. and B. M. Pötscher (2006). Can one estimate the conditional distribution of post-model-selection estimators? *Annals of Statistics*, *34*, 2554–91.

Leeb, H. and B. M. Pötscher (2008). Sparse estimators and the oracle property, or the return of Hodges' estimator. *Journal of Econometrics 142*, 201–11.

Liao, Z. (2013). Adaptive GMM shrinkage estimation with consistent moment selection. *Econometric Theory 29*, 857–904.

Lockhart, R.J., R. Taylor and R. Tibshirani (2014). A significance test for the lasso. *Annals of Statistics 42*, 413–68.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024

Luo, Y. and V. Chernozhukov (2016). Selecting informative moments via lasso, Working paper.
Milligan, K. and M. Rekkas (2008). Campaign spending limits, incumbent spending, and election outcomes, *Canadian Journal of Economics 41*, 1351–74.
Montero, S. (2016). Going it alone? An empirical study of coalition formation in elections, Working paper.
Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy 9*, 513–48.
Poole, K. T. and H. Rosenthal (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, *29*, 357–84.
Rekkas, M. (2007). The impact of campaign spending on votes in multiparty elections. *Review of Economics and Statistics 89*, 573–85.
Reynaert, M. and F. Verboven (2014). Improving the performance of random coefficients demand models: The role of optimal instruments, *Journal of Econometrics 179*, 83–98.
Rothschild, M. L. (1978). Political advertising: A neglected policy issue in marketing. *Journal of Marketing Research*, *15*, 58–71.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–88.
Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, *38*, 894–942.
Zhang, C.-H. and J. Huang (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *Annals of Statistics*, *36*, 1567–94.
Zou, H. and T. Hastie (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67*, 301–20.
Zou, H. and R. Li (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics 36*, 1509.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix
Replication package

Co-editor Victor Chernozhukov handled this manuscript.

© 2019 Royal Economic Society.

Downloaded from https://academic.oup.com/ectj/article/22/3/262/5531304 by Università Bocconi user on 01 March 2024