USING PENALIZED LIKELIHOOD TO SELECT PARAMETERS IN A RANDOM COEFFICIENTS MULTINOMIAL LOGIT MODEL

by

Joel L. Horowitz[1]
Department of Economics
Northwestern University
Evanston, IL 60201
U.S.A

and

Lars Nesheim
CeMMAP
University College London
Institute for Fiscal Studies
London, U.K.

December 2019

Abstract

This paper is about estimating a random coefficients logit model in which the distribution of each coefficient is characterized by finitely many parameters, some of which may be zero. The paper gives conditions under which, with probability approaching 1 as the sample size increases, penalized maximum likelihood (PML) estimation with the adaptive LASSO (AL) penalty distinguishes correctly between zero and non-zero parameters. The paper also gives conditions under which PML reduces the asymptotic mean-square estimation error of any continuously differentiable function of the model's parameters. The paper describes a method for computing PML estimates and presents the results of Monte Carlo experiments that illustrate their performance. It also presents the results of PML estimation of a random coefficients logit model of choice among brands of butter and margarine in the British groceries market.

Key words: Penalized estimation, adaptive LASSO, random coefficients, logit model

JEL Codes: C13, C18, C25

[1] Corresponding author

USING PENALIZED LIKELIHOOD TO SELECT PARAMETERS IN A RANDOM COEFFICIENTS A MULTINOMIAL LOGIT MODEL

1. INTRODUCTION

The multinomial logit model with random coefficients is widely used in demand modeling, empirical industrial organization, marketing, and transport economics. See, for example, Train (2009); Keane and Wasi (2013); and Ackerberg, Benkard, Berry, and Pakes (2007). Random coefficients enable taste or utility function parameters to vary among individuals in ways that are not explained by the variables available in the data. Random coefficients also enable the model to approximate any discrete choice model arbitrarily well (McFadden and Train 2000). This paper is concerned with estimating a random coefficients logit model in which the distribution of each coefficient is characterized by finitely many parameters, for example the mean and variance. Some of these parameters may be zero. The paper describes a penalized likelihood method for selecting and estimating the non-zero parameters.

In applied research, the objects of interest in a discrete choice model, such as market shares, the value of travel time, and elasticities, are smooth functions of the parameters. Some parameters, such as the mean coefficient of a price, may also be objects of interest. The mean square estimation errors of objects of interest often can be reduced by identifying and dropping from the model parameters whose values equal zero. Thus, for example, if the variance of the coefficient of a certain variable in a random coefficients logit model is zero, then the mean-square estimation errors of market shares and other objects of interest often can be reduced by treating that coefficient as a non-stochastic constant. In applications, it is not known *a priori* which parameters are zero. This paper gives conditions under which, asymptotically, penalized maximum likelihood (PML) estimation with the adaptive LASSO (AL) penalty function distinguishes correctly between zero and non-zero parameters, thereby often reducing the asymptotic mean-square estimation errors of non-zero parameters and other objects of interest in applied research.[1] We also show that the PML estimates of non-zero parameters are $n^{-1/2}$-consistent and asymptotically normally distributed, where n is the size of the estimation sample. The estimates of the non-zero parameters have the same asymptotic normal distribution that they would have if it were known *a priori* which parameters are zero and non-zero, the parameters whose values are zero were dropped from the model, and the non-zero parameters were estimated by maximum likelihood. This property is called oracle efficiency. We illustrate the numerical performance of our PML method with the results of

[1] The adaptive LASSO has certain computational advantages, including convexity of its penalty function, but the theoretical results of this paper can be obtained with other, non-convex penalty functions such as the SCAD, bridge, and minimax concave penalties. The theoretical results can also be extended to treat models with parameters that are close to but not necessarily equal to zero. See Horowitz and Huang (2013).

Monte Carlo experiments and an empirical application to choice among brands of butter and margarine in the British groceries market.

Penalization can also improve the finite-sample properties of parameter estimates in high-dimensional models when many of the parameter values are zero. In such cases, maximum likelihood (ML) estimates can have large finite-sample biases, thereby causing estimates of elasticities and other objects of substantive interest to be inaccurate. We find in our Monte Carlo experiments that the finite-sample biases of PML parameter estimates are lower, and the estimates of quantities such as elasticities are more accurate.

A further benefit of penalization is improvement in numerical accuracy. Penalized estimation with a suitable penalty function can yield parameter estimates that are true zeroes, often within a few iterations of the numerical algorithm. This is important in high-dimensional random coefficients logit models. Estimation of these models requires high-dimensional numerical integration. Dropping variance parameters that are zero and treating the associated coefficients as fixed reduces the dimension of the integral as well as the dimension of the parameter vector, thereby increasing the numerical accuracy with which the non-zero parameters are estimated. Knittel and Metaxoglu (2014) explore the numerical accuracy and consequences of numerical inaccuracy in estimation of random coefficients logit models.

This paper makes the following main contributions.

1. It shows that with probability approaching 1 as $n \to \infty$, PML estimation with the AL penalty function distinguishes correctly between zero and non-zero parameters in a random coefficients logit model. The estimates of the non-zero parameters are oracle efficient.
2. It gives conditions under which, if one or more parameters equal zero, PML estimation with the AL penalty function reduces the asymptotic mean-square errors of the estimates of non-zero parameters and often reduces the asymptotic mean square estimation error of any continuously differentiable function of the model's parameters, including predicted market shares and elasticities.
3. It describes a method for computing the PML estimates of a random coefficients logit model with the AL penalty function.
4. It presents the results of Monte Carlo experiments that illustrate the numerical performance of PML estimation of a random coefficients logit model with the AL penalty function.
5. It presents the results of PML estimation of a random coefficients logit model of choice among brands of butter and margarine in the British groceries market.

Contributions 1 and 2 above extend results of Fan and Li (2001) and Zou (2006) as well as the very large literature on penalized estimation of high-dimensional models. Fan, Lv, and Qi (2011), Horowitz (2015), and Bühlmann and van de Geer (2011) review and provide references to that literature.

Contribution 3 provides a new method to carry out PML computation that avoids the need for maximizing a non-smooth objective function and permits the use of recent advances in algorithms for solving constrained optimization problems.

The remainder of this paper is organized as follows. Section 2 describes the random coefficients logit model that we consider, PML with the AL penalty function, asymptotic properties of the parameter estimates, and asymptotic properties of smooth functions of the PML parameter estimates. Section 3 describes our method for computing the PML parameter estimates. Section 4 presents the results of the Monte Carlo experiments. Section 5 presents the application to choice among brands of butter and margarine, and Section 6 presents conclusions. The appendix presents the proofs of this paper's theoretical results.

2. THE MODEL AND ADAPTIVE LASSO ESTIMATION

Section 2.1 describes the random coefficients logit model and the penalized maximum likelihood estimation procedure that we use. Section 2.2 presents asymptotic distributional properties of the PML parameter estimates and functions of the estimates.

2.1 The Model and Estimation Procedure

Let each of n individuals choose among J exhaustive and mutually exclusive alternatives. Let $X \in \mathbb{R}^K$ denote the vector of the model's observed covariates, and let X_{ij} denote the value of X for individual i and alternative j $(j = 1,...,J)$. The indirect utility of alternative j to individual i $(i = 1,...,n)$ is

$$U_{ij} = (\beta' + \varepsilon_i')X_{ij} + \nu_{ij},$$

where ν_{ij} is a random variable with the Type I extreme value distribution, ν_{ij} and $\nu_{i'j'}$ are independent of one another if $i \neq i'$ or $j \neq j'$, β is a $K \times 1$ vector of constant coefficients, and ε_i is a $K \times 1$ vector of unobserved random variables that have means of 0 and are independently and identically distributed among individuals. In this paper, we assume that $\varepsilon_i \sim N(0_K, \Sigma)$ for each $i = 1,...,n$, where 0_K is a K-vector of zeroes and Σ is a positive-semidefinite $K \times K$ matrix. However, the paper's theoretical results hold with other distributions. Let $\phi(\xi; 0_K, \Sigma)$ denote the probability density function of the $N(0_K, \Sigma)$ distribution evaluated at the point ξ. Then the probability that individual i chooses alternative j is

$$(2.1) \qquad \pi_{ij}(\beta,\Sigma;X_{i1},...,X_{iJ}) = \int \left\{ \frac{\exp[(\beta'+\varepsilon')X_{ij}]}{\sum_{k=1}^{J}\exp[(\beta'+\varepsilon')X_{ik}]} \right\} \phi(\varepsilon;0_K,\Sigma)d\varepsilon \,.$$

Let $\Sigma = CC'$ denote the Cholesky factorization of Σ, $\tilde{\varepsilon} \sim N(0_K, I_{K\times K})$, and ϕ_K denote the $N(0_K, I_{K\times K})$ probability density function. The standard Cholesky factorization applies to full rank matrices. However, when $rank(\Sigma) = r < K$, there is a unique Cholesky factorization with $K - r$ zeroes along the diagonal of C. Therefore (2.1) can be written as

$$(2.2) \qquad \pi_{ij}(\beta,C;X_{i1},...,X_{iJ}) = \int \left\{ \frac{\exp[(\beta'+\tilde{\varepsilon}'C')X_{ij}]}{\sum_{k=1}^{J}\exp[(\beta'+\tilde{\varepsilon}'C')X_{ik}]} \right\} \phi_K(\tilde{\varepsilon})d\tilde{\varepsilon} \,.$$

The integral in (2.2) reduces to an r dimensional integral when $r < K$.

Define the choice indicator

$$d_{ij} = \begin{cases} 1 \text{ if individual } i \text{ chooses alternative } j \\ 0 \text{ otherwise} \end{cases}$$

Let $\{d_{ij}, X_{ij} : i = 1,...,n;\ j = 1,...,J\}$ be the observed choice indicators and covariates of an independent random sample of n individuals. Define $\theta = vec(\beta, C)$ and $L = \dim(\theta)$. The log-likelihood function for estimating θ is

$$\log L(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_{ij}(\theta; X_{i1},...,X_{iJ}) \,.$$

Define the maximum likelihood estimator[2]

$$\bar{\theta} = \arg\max_{\theta} \log L(\theta) \,.$$

The penalized log-likelihood function that we treat here is

$$(2.3) \qquad \log L_P(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_{ij}(\theta; X_{i1},...,X_{iJ}) - \lambda_n \sum_{\ell=1}^{L} w_\ell \,|\, \theta_\ell \,|,$$

[2] There are multiple values of θ that maximize $\log L(\theta)$. All give equal values of the maximum. Under Assumption 1 below, there is exactly one θ that maximizes $\log L(\theta)$ and for which the diagonal elements of C are non-negative. Define $\bar{\theta}$ to be this maximizer. Arguments like those used to prove Theorem 4.1.1 of Amemiya (1985) show that $\bar{\theta}$ is consistent. It is also the unique MLE subject to the constraint that the diagonal elements of C are non-negative.

where $\lambda_n > 0$ is a constant whose value may depend on n and the w_ℓ's are non-negative weights. Penalized maximum likelihood estimation with the adaptive LASSO penalty function consists of the following two steps.

Step 1: Let $\tilde{\theta}$ be a $n^{-1/2}$-consistent estimator of θ_0, possibly but not necessarily $\overline{\theta}$. Depending on how $\tilde{\theta}$ is obtained, some of its components may be zero. Define weights

$$\tilde{w}_\ell = \begin{cases} 1/|\tilde{\theta}_\ell| \text{ if } \tilde{\theta}_\ell \neq 0 \\ 0 \text{ if } \tilde{\theta}_\ell = 0. \end{cases}$$

Step 2: Let θ^* be a $L \times 1$ vector whose ℓ'th component is zero if $\tilde{\theta}_\ell = 0$ and whose remaining components are unspecified. Let $\pi_{ij}(\theta^*, X_{i1}, ..., X_{iJ})$ be the probability that individual i chooses alternative j when the parameter value is θ^*. The second-step penalized log-likelihood function is

$$\log L_P(\theta^*) = \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_{ij}(\theta^*; X_{i1}, ..., X_{iJ}) - \lambda_n \sum_{\ell=1}^{L} \tilde{w}_\ell |\theta_\ell^*|. \tag{2.4}$$

The second-step parameter estimator is

$$\hat{\theta} = \arg\max_{\theta^*} \log L_P(\theta^*),$$

where maximization is over the non-zero components of θ^*. Thus, $\hat{\theta}$ is the $L \times 1$ vector obtained by setting any parameters estimated to be 0 in the first stage equal to 0 in the π_{ij}'s, the penalty function, and $\hat{\theta}$; and maximizing the penalized log-likelihood function (2.4) over the remaining parameters. Asymptotic distributional properties of $\hat{\theta}$ and functions of $\hat{\theta}$ are described in Section 2.2.

2.2 Asymptotic Properties $\hat{\theta}$

This section describes asymptotic distributional properties of the second-step PML estimator $\hat{\theta}$ and smooth functions of $\hat{\theta}$. We assume that $\tilde{\theta} = \overline{\theta}$ and let θ_0 denote the true but unknown value of θ. Make the assumption

Assumption 1: (i) θ_0 is uniquely identified. (ii) $n^{1/2}(\overline{\theta} - \theta_0)$ converges in distribution as $n \to \infty$ to a random variable with mean 0 and covariance matrix $\overline{\Omega}$.

Let C_{jk} $(k = 1, ..., j)$ denote the components of C in the row containing C_{jj}. It is easy to show that if $C_{jj} = 0$ for some $j = 1, ..., K$, then $\partial \log L(\theta_0) / \partial C_{jk} = 0$. Therefore, $\overline{\Omega}$ can be written in the partitioned form

$$\overline{\Omega} = \begin{pmatrix} \overline{\Omega}_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\overline{\Omega}_{11}$ is the submatrix of $\overline{\Omega}$ corresponding to components of β and components C in rows for which $C_{jj} \neq 0$. We make

Assumption 1(iii): $\overline{\Omega}_{11}$ is non-singular.

Let θ_{k0} denote the k'th component of θ_0. Any parameter θ_{k0} may be zero or non-zero. Let A_0 denote the set of non-zero parameters. Its complement $\overline{A}_0$ is the set of parameters whose values are zero. Let $\hat{\theta}_k$ denote the k'th component of $\hat{\theta}$.

Assumption 2: As $n \to \infty$, $\lambda_n \to \infty$ and $n^{-1/2}\lambda_n \to 0$.

Define $\theta_{A_0} = \{\theta_{k0} : \theta_{k0} \in A_0\}$, $\theta_{\overline{A}_0} = \{\theta_{k0} : \theta_{k0} \in \overline{A}_0\}$, $\hat{\theta}_{A_0} = \{\hat{\theta}_k : \theta_{k0} \in A_0\}$, and $\hat{\theta}_{\overline{A}_0} = \{\hat{\theta}_k : \theta_{k0} \in \overline{A}_0\}$. Let $\overline{\theta}_{A_0}$ be the unpenalized ML estimator of θ_{A_0} in model (2.1) when one correctly fixes $\theta_{\overline{A}_0}$ equal to the zero vector.

Assumption 3: As $n \to \infty$, $n^{1/2}(\overline{\theta}_{A_0} - \theta_{A_0}) \to^d N(0, \Omega_{A_0})$ for some non-singular covariance matrix Ω_{A_0}.

Amemiya (1985), among many others, gives primitive conditions for assumption 3.

Let $g(\theta)$ ($\theta \in \mathbb{R}^L$) be a continuously differentiable function. The PML estimate of $g(\theta_0)$ is $g(\hat{\theta})$. The unpenalized ML estimate is $g(\overline{\theta})$. Let $AMSE[g(\hat{\theta})]$ and $AMSE[g(\overline{\theta})]$, respectively, denote the asymptotic mean square errors (AMSE's) of $g(\hat{\theta})$ and $g(\overline{\theta})$ as estimators of $g(\theta_0)$. Let $I_{full} = \overline{\Omega}_{11}^{-1}$ denote the information matrix of the version of model (2.2) (that is $-E \log \partial^2 L(\theta_0) / \partial\theta\partial\theta'$) that is obtained by setting $C_{jk} = 0$ ($k = 1, ..., j$) deterministically if $C_{jj} = 0$. Partition I_{full} as

$$I_{full} = \begin{pmatrix} I_{11} & I_{12} \\ I_{12}' & I_{22} \end{pmatrix},$$

where $I_{11} = -E[n^{-1}\partial^2 \log L(\theta_0) / \partial\theta_{A_0}\partial\theta_{A_0}']$ is the submatrix of I_{full} corresponding to θ_{A_0}, $I_{22} = -E[n^{-1}\partial^2 \log L(\theta_0) / \partial\theta_{\overline{A}_0}\partial\theta_{\overline{A}_0}']$ is the submatrix of I_{full} corresponding to components of $\theta_{\overline{A}_0}$ other than components C_{jk} for which $C_{jj} = 0$, and $I_{12} = -E[n^{-1}\partial^2 \log L(\theta_0) / \partial\theta_{A_0}\partial\theta_{\overline{A}_0}']$ is the submatrix of I_{full} corresponding to the covariance of the estimators of θ_{A_0} and the components of $\theta_{\overline{A}_0}$ just described.

Let $\overline{\Omega}_{A_0}$ denote the submatrix of I_{full}^{-1} corresponding to the components of θ_{A_0} The following theorem gives the main theoretical result of this paper.

Theorem 2.1: Let assumptions 1-3 hold. As $n \to \infty$

(i) $P(\hat{\theta}_k = 0 \; \forall \; k \text{ such that } \theta_{k0} \in \overline{A}_0) \to 1$

(ii) $n^{1/2}(\hat{\theta}_{A_0} - \theta_{A_0}) \to^d N(0, \Omega_{A_0})$

(iii) If $\overline{A}_0$ is non-empty, then $\overline{\Omega}_{A_0} - \Omega_{A_0}$ is positive semidefinite and positive definite if I_{12} has full rank.

(iv) If $\overline{A}_0$ is non-empty, then $AMSE[g(\overline{\theta})] \geq AMSE[g(\hat{\theta})]$.

Parts (i) and (ii) of Theorem 2.1 state that PML estimation with the AL penalty function distinguishes correctly between zero and non-zero parameters with probability approaching 1 as $n \to \infty$. Part (ii) states that the PML estimates of the non-zero parameters are oracle efficient. That is, they have the same asymptotic normal distribution that they would have if it were known which parameters in model (2.1) are zero and non-zero, the parameters whose values are zero were dropped from the model, and the non-zero parameters were estimated by maximum likelihood. Part(iii) implies that that if one or more parameters are zero and I_{12} has full rank, then PML estimation with the AL penalty function reduces the asymptotic mean-square estimation errors of the model's non-zero parameters. Part (iv) states that the asymptotic mean square error of the unpenalized ML estimate of $g(\theta_0)$ is at least as large as the asymptotic mean square error of the PML estimate. Section 4 gives examples in which the mean square error of $g(\overline{\theta})$ is much greater than the mean square error of $g(\hat{\theta})$. Asymptotic inference about $g(\theta_0)$ based on PML estimation can be carried out by applying the delta method to $g(\hat{\theta})$.

3. COMPUTATION

Maximizing $\log L_P(\theta)$ presents several computational problems. There may be more than one local maximum of $\log L_P(\theta)$, the penalty function in $\log L_P(\theta)$ is not differentiable at all values of θ, and $\log L_P(\theta)$ includes high-dimensional integrals that must be evaluated numerically. We deal with the first of these problems by maximizing $\log L_P(\theta)$ repeatedly using a different initial value of θ each time.

We deal with the second by reformulating the optimization problem to one of maximizing a differentiable objective function subject to linear constraints. To do this, write $\theta = \theta^+ - \theta^-$, where θ^+

and θ^- are $L \times 1$ vectors whose components are non-negative. Then maximizing $\log L_P(\theta)$ in (2.3) is equivalent to solving the problem

$$(3.1) \qquad \underset{\theta, \theta^+, \theta^-}{\text{maximize}}: \sum_{i=1}^{n}\sum_{j=1}^{J} d_{ij} \log \pi_{ij}(\theta; X_{i1}, ..., X_{iJ}) - \lambda_n \sum_{\ell=1}^{L} w_\ell (\theta_\ell^+ + \theta_\ell^-)$$

subject to

$$\theta = \theta^+ - \theta^-$$
$$\theta^+, \theta^- \geq 0,$$

where the last inequality holds component by component. This formulation avoids the need to maximize a non-smooth objective function and permits exploitation of advances in methods for solution of constrained optimization problems.

There is a large econometric literature on numerical methods for evaluating high-dimensional integrals. See, for example, McFadden (1989); McFadden and Ruud (1994); Geweke, Keane, and Runkle (1994); Hajivassiliou, McFadden, and Ruud (1996); Geweke and Keane (2001), and Train (2009). Available methods include Gaussian integration procedures, pseudo Monte Carlo procedures, quasi Monte Carlo procedures, and Markov chain Monte Carlo (MCMC) methods. More recently, Heiss and Winschel (2008), Skrainka and Judd (2011), and Knittel and Metaxoglou (2014) have suggested that sparse grid integration methods produce accurate approximations at low cost. To focus on the performance of the PML method, we have used a simple pseudo Monte Carlo integration method based on either 500 or 1500 draws from a normal random number generator.

We computed the solution to problem (3.1) by using a sequential quadratic programming algorithm for constrained optimization from the NAG Fortran Library (The Numerical Algorithms Group, Oxford U.K., www.nag.com). The algorithm is based on NPOPT, which is part of the SNOPT package described by Gill, Murray, and Saunders (2005).

4. MONTE CARLO EXPERIMENTS

This section reports the results of a Monte Carlo investigation of the numerical performance of the PML method. We used two designs. One is based on a small, hypothetical model. The other is based on data from the U.K. market for butter and margarine.

4.1 Design 1: A Hypothetical Model

This design consists of a model with $J = 5$ alternatives in the choice set and $K = 20$ covariates. The random coefficients are independent of one another, so their covariance matrix is diagonal. The means and variances of the coefficients are as follows:

k	Mean (β_k)	Variance $Var(\varepsilon_k)$
$1 \le k \le 2$	1	1
$3 \le k \le 5$	1	0
$6 \le k \le 20$	0	0

Thus, there are two non-zero random coefficients, three non-zero coefficients that are not random, and 15 non-random coefficients whose values are zero. The covariates are independently distributed as $N(0,1)$. The sample size is $n = 1000$.

We carried out PML estimation with 300 simulated datasets and chose the penalty parameter λ_n to minimize the Bayes Information Criterion (BIC) using the computational procedure described in the next paragraph. Wang, Li, and Tsai (2007) and Wang, Li, and Leng (2009) give conditions under which penalized estimation of a linear model is model-selection consistent when the penalty parameter is chosen by the BIC. The theoretical properties of the BIC for PML estimation have not been studied. We used a pseudo Monte Carlo numerical integration procedure with antithetic variates with 500 draws from a 10-dimensional random number generator. We assumed that only 10 covariates, including the first 5, have potentially non-zero variances. Therefore, 30 parameters were estimated.

We chose λ_n by solving (2.3) for the two steps of the adaptive LASSO procedure using each point in a rectangular grid of values of λ_n. There were 5 grid points for step one of the adaptive LASSO procedure, 10 points for step 2, and 50 points in total. The values of the step 1 points ranged from 10^{-4} to 10^{-3}. The values of the step 2 points ranged from 10^{-4} to 10^{-2}. The logarithms of the values in each dimension of the grid were equally spaced. We report results for the grid point of λ_n values that minimizes the BIC in step 2.

The time required to compute the PML estimates depends on the number of grid points used for selection of the penalty parameters by the BIC. For Design 1, the time is trivial. For Design 2, which is discussed in detail in Section 4.2, we found the PML and ML computation times in the Monte Carlo experiment to be related by

$$t_{PML} \approx (0.96 n_1 + 0.46 n_2) t_{ML},$$

where t_{PML} and t_{ML}, respectively, are the PML and ML computation times, and n_1 and n_2, respectively, are the numbers of grid points used to select the first- and second-step penalty parameters. The first stage of the adaptive LASSO procedure has approximately the same computation time as unpenalized ML. The second stage has approximately 46% of the computation time of the unpenalized ML. This is because the first stage provides a good set of initial parameter values and sets many equal to zero, thereby reducing

the number of iterations required for convergence of the algorithm described in Section 3. Given values of the penalty parameters, the combined computation time of the two stages of adaptive LASSO estimation is about 42% longer than the computation time of unpenalized ML estimation.

We investigated the sensitivity of the mean-square errors (MSEs) of the PML estimates to variations in the values of the penalty parameters around the values selected by the BIC. We found that increasing the values by 20% had little effect on the MSE's, but decreasing the values by 20% caused large increases in the MSE's. We obtained similar results in the Monte Carlo experiment reported in Section 4.2. We conclude that the BIC selects the penalty parameters satisfactorily but do not claim that the BIC makes selections that are optimal in any sense.

The results of the experiment are shown in Table 1. The average number of non-zero parameters in the model estimated by PML is 7.22, compared to 30 potentially non-zero parameters in the full model. With probability 1, unconstrained maximum likelihood estimation cannot yield estimates of zero, so unconstrained maximum likelihood estimation gives 30 non-zero parameter estimates. The MSE's of the PML estimates of the means of the non-zero slope coefficients (the non-zero β_k's) are all less than 33% and often less than 10% of the MSE's of the unconstrained ML estimates. The MSE's of the PML estimates of the standard deviations are approximately 10% of the MSE's of the unconstrained maximum likelihood estimates. In summary, PML selects a smaller model and gives estimates of important parameters with much smaller mean-square errors than does unconstrained maximum likelihood estimation.

4.2 Design 2: Butter and Margarine

This design is based on data about the U.K. market for butter and margarine. The data were obtained by the research company Kantar and used by Griffith, Nesheim, and O'Connell (2015). The data contain information on 10,102 households that shopped at supermarkets in the U.K. The data include demographic characteristics of each household (e.g., household size, age, employment status, and average weekly grocery expenditure), product characteristics (e.g., brand, package size, and saturated fat content), and consumer purchase choices. On each shopping trip, each consumer chose either not to buy any product or to buy one of 47 products available in the market. Thus, the number of options in each consumer's choice set is $J = 48$.

The Kantar data contain $K = 50$ covariates, including product fixed effects. Thus, the choice model of equation (2.2) contains 99 parameters. There are 49 mean parameters (the components of β in (2.2)) and 50 variance parameters. The mean parameter for the outside option of no purchase is normalized to be zero. In the Monte Carlo experiment, we set the parameters equal to the penalized maximum likelihood estimates obtained from a random sample of 5000 observations from the Kantar

data. The resulting model (the "true model") has 37 non-zero mean parameters and four non-zero random coefficient variance parameters. The remaining 58 parameters of the true model are zero. We used this model to simulate the product choices of 5000 hypothetical households. We used the simulated choice data to estimate the choice model's 99 parameters using unpenalized ML, PML, and the oracle ML (maximum likelihood estimation of only the 41 non-zero parameters of the true model and the remaining parameters set equal to zero). We used 1500 antithetic variate draws from a multivariate normal random number generator to compute the numerical integral.

Table 2 summarizes results of 300 Monte Carlo replications of the foregoing simulation procedure. Columns 3-5 show the MSEs of the estimates of the non-zero parameters of the true model using each estimation method. The parameter β_1 is the mean price coefficient in the model. The MSEs of the PML estimates of 40 of the 41 parameters are significantly smaller than those of the unpenalized ML estimates and closer to the MSEs of the oracle ML estimates. The MSEs of the PML estimates of 27 parameters are less than 10% of the unpenalized ML estimates. The MSEs of 8 PML estimates are 10%-20% less than the unpenalized MSEs, and the MSEs of 5 PML estimates are 21%-47% less than the unpenalized MSEs. For example, the MSE of the PML estimate of β_1 is 0.114 compared to 1.50 for the unpenalized ML estimate and 0.104 for the oracle ML estimate. The mean number of non-zero parameters in the selected model is 31.8, and 90 percent of the replications select a model with 29-37 non-zero parameters. The coefficient of the price variable is non-zero in all replications.

We also computed the own-price elasticities of the 47 products (excluding no-purchase option) in each Monte Carlo replication. The MSE's of 33 of the 47 elasticity estimates obtained by PML were less than the MSE's of the corresponding elasticity estimates obtained by unpenalized ML. The median ratio of the MSEs of the unpenalized ML and PML elasticity estimates is 1.49. That is the median value of (MSE of ML estimates)/(MSE of PML estimates) is 1.49. The median ratio of the MSEs of the PML and oracle ML elasticity estimates is 0.995. Thus, the PML elasticity estimates, like the PML parameter estimates, are more accurate than the estimates obtained from unpenalized ML and close to the oracle estimates.

The biases of the ML coefficient and elasticity estimates are larger than the biases of the corresponding PML estimates. This is illustrated by Figures 1 and 2, which are histograms of the empirical distributions of estimates obtained through Monte Carlo replication. Figure 1 shows the differences between the ML (blue) and PML (orange) estimates of the price coefficient and true value of this coefficient. Figure 2 shows the differences between the ML (blue) and PML (orange) estimates of own price elasticities and true values. The figures show that the unpenalized ML estimates have large biases and are highly variable. The difference between the ML estimates and true value of the price coefficient has a mode at -1.12, compared to a mode of -0.35 for the PML estimates. The difference

between the ML estimates and true values of the own-price elasticity has a mode at 0.68, compared to -0.52 for the PML estimates. The locations of these modes are finite-sample effects. The asymptotic distributions of the ML and PML estimates are both centered at zero. The reduction of finite-sample bias through PML is due to the reduction in the number of non-zero parameters that must be estimated. This is an important advantage of PML.

To illustrate the performance of PML in policy analysis, we used the PML, unpenalized ML, and oracle estimates to evaluate effects of a 20% value added tax (VAT) on butter and margarine. Currently, food purchases in the UK are exempt from the VAT. The VAT increases the prices of butter and margarine, which reduces demand for these products, consumer welfare, and revenues from the sale of butter and margarine. We computed four resulting economic effects. The first is the reduction in consumer welfare as measured by the compensating income variation for the VAT. The second is the reduction in revenues to sellers of butter and margarine. The third is tax revenues resulting from the VAT. The fourth is the changes in the market shares of the products. We assumed that the pre-tax prices of butter, margarine, and any substitute products remain unchanged.

We now describe how we computed the foregoing effects. Let X_j^{notax} denote the values of the explanatory variables for product j in model (2.2) before the VAT and X_j^{tax} denote the values of the same variables after the prices of butter and margarine have been increased by 20%. Let p_j denote the before-VAT price of product j, τ denote the tax rate, and $p_j^{tax} = (1+\tau)p_j$ denote the price after the VAT has been imposed. Denote the mean and random component of the coefficient of price in (2.2) by β_1 and $\tilde{\varepsilon}_1 C_{11}$, respectively. The consumer compensating variation for the tax increase is (Small and Rosen 1981)

$$CV(\beta, C) = \sum_{i=1}^{5000} \int \frac{\log\left[\sum_{j=0}^{47} \exp(\beta + \tilde{\varepsilon}' C') X_{ij}^{notax}\right] - \log\left[\sum_{j=0}^{47} \exp(\beta + \tilde{\varepsilon}' C') X_{ij}^{tax}\right]}{\beta_1 + \tilde{\varepsilon}_1 C_{11}} \phi(\tilde{\varepsilon}) d\tilde{\varepsilon} .$$

The change in revenues is

$$\Delta R = \sum_{j=1}^{47} \sum_{i=1}^{5000} p_j [\pi_{ij}(\beta, \Sigma; X_j^{tax}) - \pi_{ij}(\beta, \Sigma; X_j^{notax})] .$$

The change in the market share of product j is

$$\Delta S_j = \sum_{i=1}^{5000} [\pi_{ij}(\beta, \Sigma; X_j^{tax}) - \pi_{ij}(\beta, \Sigma; X_j^{notax})] .$$

ΔR is the change the revenues of sellers after remitting tax revenues of τR^{tax} to the government and, therefore, does not include the factor $1+\tau$. The sums are over the 47 products and 5000 individuals in the experiment.

Table 3 shows the MSEs of the estimated effects of the VAT. The table shows the MSEs of the estimated change in median market share, not the MSEs of the estimated changes in the shares of individual products. The MSEs of the unpenalized ML and PML estimates of the compensating variation are similar. The MSEs of the PML estimates of the change in revenues to sellers (in pounds per trip per individual), tax revenues, and change in median market share are smaller than the MSEs of the unpenalized ML estimates of these quantities and closer to the oracle estimates.

5. EMPIRICAL APPLICATION

This section summarizes the results of applying the PML and unpenalized ML methods to the full Kantar data set that is described in the first paragraph of Section 4.2. We compare the own price elasticities obtained with the two methods and the results of the tax experiment described in Section 4.2. As is explained in the second paragraph of Section 4.2, the model has 99 parameters, including 49 means of the random slope coefficients and 50 standard deviations. All of the unpenalized parameter estimates are non-zero, and the empirical Hessian matrix has full rank. Only 35 of the penalized estimates are non-zero, including 31 slope coefficients and 4 standard deviation parameters.

Table 4 shows summary statistics for own price elasticities. On average the PML and unpenalized ML elasticity estimates are similar in magnitude, but the PML estimates are less dispersed.

Table 5 shows summary statistics for changes in market shares and product revenues in (in units of pounds per shopping trip per individual) the tax experiment. The mean change in market share is zero because the sum of the shares must equal one. The PML estimate of the change in market shares is more dispersed than the unpenalized ML estimate. The means and standard deviations of the PML and unpenalized ML estimates of the change in revenues are roughly equal.

The PML and unpenalized ML estimates of the compensating variation for the tax increase are 0.336 and 0.340 pounds per shopping trip, respectively. The PML and ML estimates of tax revenue, respectively, are 0.0867 and 0.0888 pounds per shopping trip. Thus, the PML estimates of the compensating variation and tax revenues are virtually identical.

6. CONCLUSIONS

This paper has been concerned with estimating a random coefficients logit model in which the distribution of each coefficient is characterized by finitely many parameters. Some of these parameters may be zero. The paper has given conditions under which with probability approaching one as the sample size approaches infinity, penalized maximum likelihood (PML) estimation with the adaptive LASSO

(AL) penalty function distinguishes correctly between zero and non-zero parameters in a random coefficients logit model. The estimates of the large parameters are oracle efficient. If one or more parameters are zero, then PML estimation with the AL penalty function often reduces the asymptotic mean-square estimation error of any continuously differentiable function of the model's parameters, such as a predicted market share. The paper has described a method for computing the PML estimates of a random coefficients logit model. It has presented the results of Monte Carlo experiments that illustrate the numerical performance of PML estimates. The paper has also presented the results of PML estimation of a random coefficients logit model of choice among brands of butter and margarine in a British grocery chain.

The Monte Carlo results show that PML estimates have lower mean-square errors and finite-sample biases than unpenalized ML estimates with sample sizes similar to those used in marketing and empirical industrial organization. PML estimation is tractable computationally, and the PML method can be modified easily for use in generalized method of moments estimation.

APPENDIX PROOF OF THEOREM 2.1

Parts (i) and (ii): Let $I^*_{full} = -E\partial^2 \log L(\theta_0)/\partial\theta\partial\theta'$ denote the information matrix of model (2.2). Define the vector u by

$$\theta = \theta_0 + n^{-1/2}u$$

for any θ. Let θ_{A_0} be the first $L_0 \equiv \dim(A_0)$ components of θ_0 and $\theta_{\bar{A}_0}$ be the remaining $L - L_0$ components. Order the components of u similarly. Define

$$\begin{aligned} D_n(u) &= \log L(\theta_0 + n^{-1/2}u) - \log L(\theta_0) \\ &\quad + \lambda_n \left[\sum_{\ell=1}^{L_0} |\tilde{\theta}_\ell|^{-1} (|\theta_{\ell 0} + n^{-1/2}u_\ell| - |\theta_{\ell 0}|) + \sum_{\ell=L_0+1}^{L} \tilde{w}_\ell |n^{-1/2}u_\ell| \right] \\ &= n^{-1/2} \frac{\partial \log L(\theta_0)}{\partial \theta'} u - (1/2)u' I^*_{full} u[1 + o_p(1)] \\ &\quad + \lambda_n \left[\sum_{\ell=1}^{L_0} |\tilde{\theta}_\ell|^{-1} (|\theta_{\ell 0} + n^{-1/2}u_\ell| - |\theta_{\ell 0}|) + \sum_{\ell=L_0+1}^{L} \tilde{w}_\ell |n^{-1/2}u_\ell| \right]. \end{aligned}$$

Write the penalty term above as

$$n^{-1/2}\lambda_n \left[\sum_{\ell=1}^{L_0} |\tilde{\theta}_\ell|^{-1} n^{1/2} (|\theta_{\ell 0} + n^{-1/2}u_\ell| - |\theta_{\ell 0}|) + \sum_{\ell=L_0+1}^{L} \tilde{w}_\ell |u_\ell| \right]$$

If $\theta_{\ell 0} \neq 0$, then

$$n^{1/2}(|\theta_{\ell 0} + n^{-1/2}u_\ell| - |\theta_{\ell 0}|) \to^p u_\ell \operatorname{sgn}(\theta_{\ell 0}),$$

where $\operatorname{sgn}(v)$ for any scalar v equals 1, -1, or 0 according to whether v is positive, negative, or zero (Zou 2006). Therefore, the terms of the penalty function corresponding to components of θ_{A_0} converge in probability to 0. The terms in the penalty function corresponding to $\theta_{\bar{A}_0}$ diverge to ∞ if $|u_\ell| > 0$ for any $\ell = L_0 + 1, ..., L$. Therefore, if the components of u corresponding to $\theta_{\bar{A}_0}$ are non-zero, D_n is dominated by the penalty term, which increases without bound as $n \to \infty$. Arguments identical to those of Zou (2006, proof of Theorem 2) except with the least-squares objective function replaced by $-\log L(\theta)$, show that if $\theta_{k0} \in \bar{A}_0$, then $P(\hat{\theta}_k \in \bar{A}_0) \to 1$. D_n is dominated asymptotically by $\log L(\theta_{A_0} + n^{-1/2}u_0, 0) - \log L(\theta_{A_0}, 0)$, where u_0 denotes the components of u corresponding to components of θ_{A_0} and the argument 0 corresponds to $\theta_{\bar{A}_0}$. Therefore, standard results for maximum likelihood estimates yield parts (i) and (ii).

Part (iii): By definition

$$I_{full} = \begin{pmatrix} I_{11} & I_{12} \\ I_{12}' & I_{22} \end{pmatrix}.$$

where I_{11} is non-singular. $I_{22} - I_{12}'I_{11}^{-1}I_{12}$ is the Schur complement of I_{11} and is non-singular because I_{full} and I_{11} are non-singular. Let $(I_{full}^{-1})_{A_0}$ denote the submatrix of I_{full}^{-1} corresponding to the components of θ_{A_0}. Then

$$\bar{\Omega}_{A_0} = (I_{full}^{-1})_{A_0} = I_{11}^{-1} + I_{11}^{-1}I_{12}(I_{22} - I_{12}'I_{11}^{-1}I_{12})^{-1}I_{12}'I_{11}^{-1} \geq I_{11}^{-1} = \Omega_{A_0}.$$

The inequality is strict if I_{12} has full rank.

Part (iv): Define $S = I_{22} - I_{12}'I_{11}^{-1}I_{12}$. Then

$$I_{full}^{-1} = \begin{pmatrix} I_{11}^{-1} + I_{11}^{-1}I_{12}S^{-1}I_{12}'I_{11}^{-1} & -I_{11}^{-1}I_{12}S^{-1} \\ -S^{-1}I_{12}'I_{11}^{-1} & S^{-1} \end{pmatrix}.$$

Moreover,

$$AMSE[g(\hat{\theta})] - AMSE[g(\bar{\theta})] = \frac{\partial g(\theta_0)}{\partial \theta'} A \frac{\partial g(\theta_0)}{\partial \theta},$$

where

$$A = \begin{pmatrix} I_{11}^{-1} I_{12} S^{-1} I_{12}' I_{11}^{-1} & -I_{11}^{-1} I_{12} S^{-1} \\ -S^{-1} I_{12}' I_{11}^{-1} & S^{-1} \end{pmatrix}.$$

The matrix A is positive semidefinite (Bekker 1988). Q.E.D.

ACKNOWLEDGEMENTS

Research carried out in part while Joel L. Horowitz was a visitor to the Department of Economics, University College London. We gratefully acknowledge financial support from the Economic and Social Research Council (ESRC) through the ESRC Centre for Microdata Methods and Practice (CeMMAP) grant number ES/1034021/1 and ESRC Large Research Grant ES/P008909/1. Data were provided by Kantar UK Ltd. The use of the data in this research does not imply endorsement by Kantar UL Ltd. of the interpretation or analysis of the data.

REFERENCES

Ackerberg, D.C., L. Benkard, S. Berry, and A. Pakes (2007). Econometric tools for analyzing market outcomes. Handbook of Econometrics, Vol. 6, ed. by E.E. Leamer and J.J. Heckman. Amsterdam: North-Holland, pp. 4171-4276.

Bekker, P.A. (1988). The positive semidefiniteness of partitioned matrices. Linear Algebra and Its Applications, 111, 261-278.

Bühlmann, P. and S. van de Geer (2011). Statistics for High-Dimensional Data. Heidelberg: Springer-Verlag.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96, 1348-1360.

Fan, J., J. Lv, and L. Qi (2011). Sparse high-dimensional models in economics. Annual Review of Economics, 3, 291-317.

Genz, A. and K.-S. Kwong (2000). Numerical evaluation of singular multivariate normal distributions. Journal of Statistical Computation and Simulation, 68, 1-21.

Geweke, J. and M. Keane (2001). Computationally intensive methods for integration in econometrics. Handbook of Econometrics, Vol. 5, ed. by E.E. Leamer and J.J. Heckman. Amsterdam: North-Holland, pp. 3463-3568.

Geweke, J., M. Keane, and D. Runkle (1994). Alternative computational approaches to inference in the multinomial probit model. Review of Economics and Statistics, 76, 609-632.

Gill, P.E., W. Murray, and M.A. Saunders (2005). Users' guide for SNOPT 7.1: a Fortran package for large-scale linear and nonlinear programming. Report NA 05-2, Department of Mathematics, University of California, San Diego. http://www.ccom.uscd.edu/~peg/papers/sndoc7.pdf.

Griffith, R. L. Nesheim, and M. O'Connell (2015). Income effects and the welfare consequences of tax in differentiated product oligopoly. Working paper, Institute for Fiscal Studies, London, U.K.

Hajivassiliou, V., McFadden, D. and P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. Journal of Econometrics, 72, 85-134.

Heiss, F. and V. Winschel (2008). Likelihood approximation by numerical integration on sparse grids. Journal of Econometrics, 144, 62-80.

Horowitz, J.L. (2015). Variable Selection and Estimation in High-Dimensional Models. Canadian Journal of Economics, 48, 389-407.

Horowitz, J.L. and J. Huang (2013). Penalized Estimation of High-Dimensional Models under a Generalized Sparsity Condition. Statistica Sinica, 23, 725-748.

Keane, M. and N. Wasi (2013). Comparing alternative models of heterogeneity in consumer choice behavior. Journal of Applied Econometrics, 28, 1018-1045.

Knittel, C.R. and K. Metaxoglou (2014). Estimation of random-coefficent demand models: Two empiricists' perspective. Review of Economics and Statistics, 96, 34-59.

McFadden, D. 1989). Simulation and the asymptotics of optimization estimators. Econometrica, 57, 1027-1057.

McFadden, D. and P.A. Ruud (1994). Estimation by simulation. Review of Economics and Statistics, 76, 591-608.

McFadden, D. and K. Train (2000). Mixed MNL models for discrete response. Journal of Applied Econometrics, 15, 447-470.

Skrainka, B.S. and K.L. Judd (2011). High performance quadrature rules: How numerical integration affects a popular model of product differentiation. Cemmap working paper CWP03/11, Institute for Fiscal Studies, London, U.K.

Train, K.E. (2009). Discrete Choice Methods with Simulation. Cambridge, U.K.: Cambridge University Press.

Wang, H., R. Li, and C.-L. Tsai (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 94, 553-568.

Wang, H., B. Li, and C. Leng (2009). Shrinkage tuning parameter selection with a diverging number of parameters. Journal of the Royal Statistical Society, Series B, 71, 671-683.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101, 1418-1429.

Table 1: Results of Monte Carlo Experiments with Design 1[a]

Parameter	MSE of PML Estimate	MSE of Unpenalized ML Estimate	MSE of Oracle ML Estimate
β_1	0.0099	0.294	0.00910
β_2	0.0103	0.333	0.00958
β_3	0.00664	0.290	0.00596
β_4	0.00672	0.303	0.00611
β_5	0.00691	0.287	0.00654
σ_1	0.0350	0.339	0.0228
σ_2	0.0294	0.313	0.0238
Average number of non-zero parameters in the model selected by PMLE	7.22		
Average value of λ in step 2	0.00835		

a. Based on 300 Monte Carlo replications. σ_1 and σ_2, respectively, are the standard deviations of ε_1 and ε_2. The correct model is the model specified in design 1 with the parameter values specified in that design. The model selected by PML contains the correct model if the PML estimates of the non-zero parameters of the correct model are not zero.

Table 2: Results of Monte Carlo Experiments with Design 2[b]

Parameter	Definition of variable	MSE of PML Estimate	MSE of Unpenalized ML Estimate	MSE of Oracle ML Estimate
β_1	Price	0.114	1.50	0.104
β_2	Index of monthly advertising expenditure	0.00235	0.0285	0.00189
β_3	Square of index of monthly advertising expenditure	0.000443	0.0506	0.000781
β_4	Dummy variable equal to 1 for 500 gram pack and 0 otherwise	1.38	14.9	0.385
β_5	Dummy variable equal to 1 for 1000 gram pack and 0 otherwise	12.2	33.2	6.81
β_6	$\beta_5 \times$ Household size	0.00274	0.0461	0.00328
β_7	Grams of saturated fat per pack	0.106	0.942	0.0425
β_8	Dummy variable equal to 1 if household size is 2 and makes no purchase and 0 otherwise	0.345	0.736	0.0189
β_9	Dummy variable equal to 1 if household size is 3 and makes no purchase and 0 otherwise	0.154	1.06	0.0537
β_{10}	Dummy variable equal to 1 if household size is 4 and makes no purchase and 0 otherwise	0.153	3.88	0.0328
β_{11}	Brand-specific constant	0.282	16.1	0.0867
β_{12}	Brand-specific constant	0.996	18.6	0.670
β_{13}	Brand-specific constant	5.18	40.5	2.11
β_{14}	Brand-specific constant	3.70	10.9	1.42
β_{15}	Brand-specific constant	1.41	68.1	0.427
β_{16}	Brand-specific constant	1.62	6.15	0.463
β_{17}	Brand-specific constant	0.147	33.4	0.0373
β_{18}	Brand-specific constant	1.61	23.4	0.904
β_{19}	Brand-specific constant	0.809	88.2	0.377
β_{20}	Brand-specific constant	4.09	23.6	1.21
β_{21}	Brand-specific constant	5.44	177	1.95
β_{22}	Brand-specific constant	5.62	26.0	1.75
β_{23}	Brand-specific constant	3.77	67.5	2.76

Table 2, continued

β_{24}	Brand-specific constant	4.50	34.5	1.48
β_{25}	Brand-specific constant	2.23	14.7	0.651
β_{26}	Brand-specific constant	3.93	65.5	1.514
β_{27}	Brand-specific constant	0.200	26.1	0.0375
β_{28}	Brand-specific constant	0.0890	45.5	0.0477
β_{29}	Brand-specific constant	0.734	8.71	0.526
β_{30}	Brand-specific constant	0.220	21.8	0.0512
β_{31}	Brand-specific constant	0.173	78.8	0.0365
β_{32}	Brand-specific constant	2.77	35.7	0.869
β_{33}	Brand-specific constant	0.567	61.3	0.211
β_{34}	Brand-specific constant	2.30	39.1	0.563
β_{35}	Brand-specific constant	3.59	196.4	1.671
β_{36}	Brand-specific constant	3.03	57.7	1.09
β_{37}	Brand-specific constant	0.626	132.0	0.234
σ_1	Standard deviation of coefficient of price	0.270	1.44	0.204
σ_6	Standard deviation of coefficient of saturated fat per pack	0.0578	1.39	0.0706
σ_{23}	Standard deviation of coefficient of a brand-specific constant	17.3	14.2	9.55
σ_{38}	Standard deviation of utility of no-purchase option for households of at least 5 persons	9.10	53.9	6.68
Average number of non-zero parameters in the model selected by PMLE		31.8		
Average value of λ in step 2		0.0035		

Based on 300 Monte Carlo replications.

Table 3: Mean Square Errors of Estimated Effects of the VAT in Monte Carlo Design 2

Effect	MSE Using Unpenalized ML	MSE Using PML	MSE Using Oracle Model
Compensating Variation	0.0150	0.0143	0.00827
Change in Revenues to Sellers	0.0259	0.0148	0.0122
Tax Revenues	0.0188	0.0170	0.0100
MSE of Change in Median Market Share	4.08×10^{-7}	3.74×10^{-7}	1.08×10^{-7}

TABLE 4: SUMMARY STATISTICS FOR OWN PRICE ELASTICITIES

Method	Mean Elasticity	Standard Deviation of Elasticity	Minimum	Maximum
MLE	-2.806	0.9159	-4.637	-1.422
PMLE	-2.713	0.6297	-4.447	-1.489

TABLE 5: SUMMARY STATISTICS FOR CHANGES IN MARKET SHARES AND PRODUCT REVENUE

Method	Standard Deviation of Change in Share ($\times 10^{-3}$)	Mean Change in Revenue $\times 10^{-1}$	Standard Deviation of Change in Revenue $\times 10^{-3}$
MLE	4.673	-4.859	7.320
PMLE	5.904	-4.866	7.338

FIGURE 1

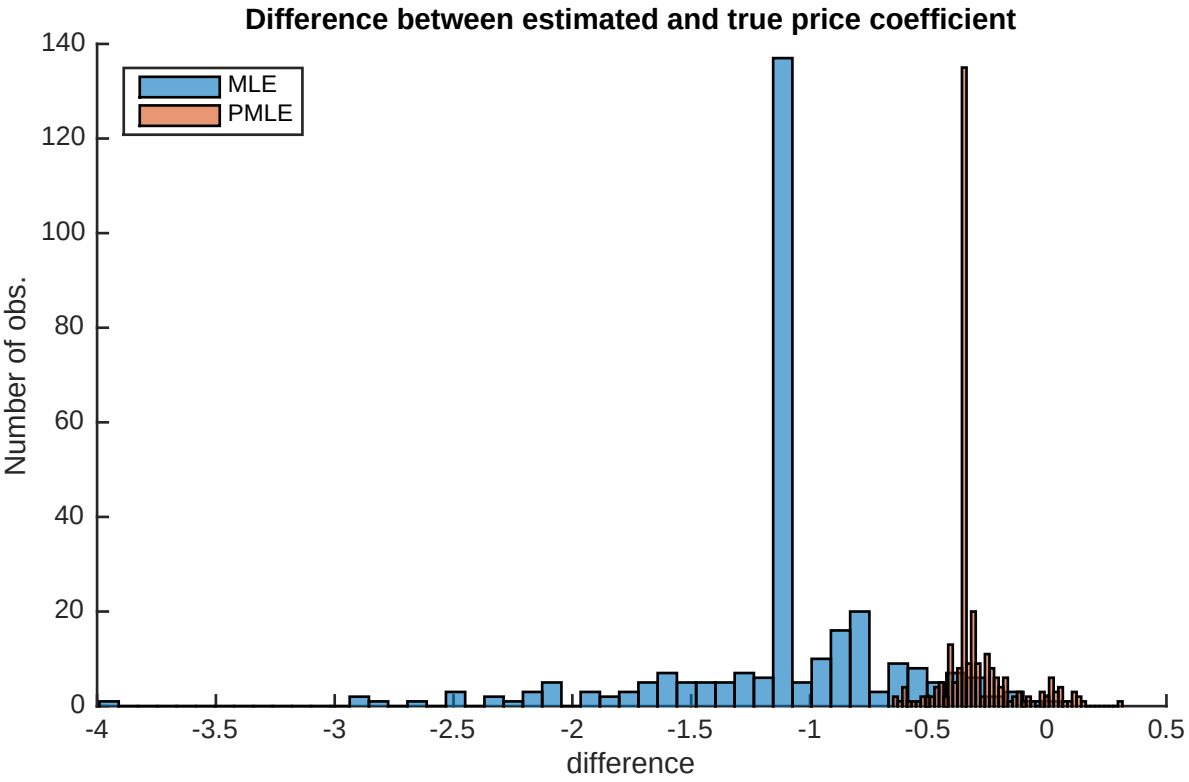
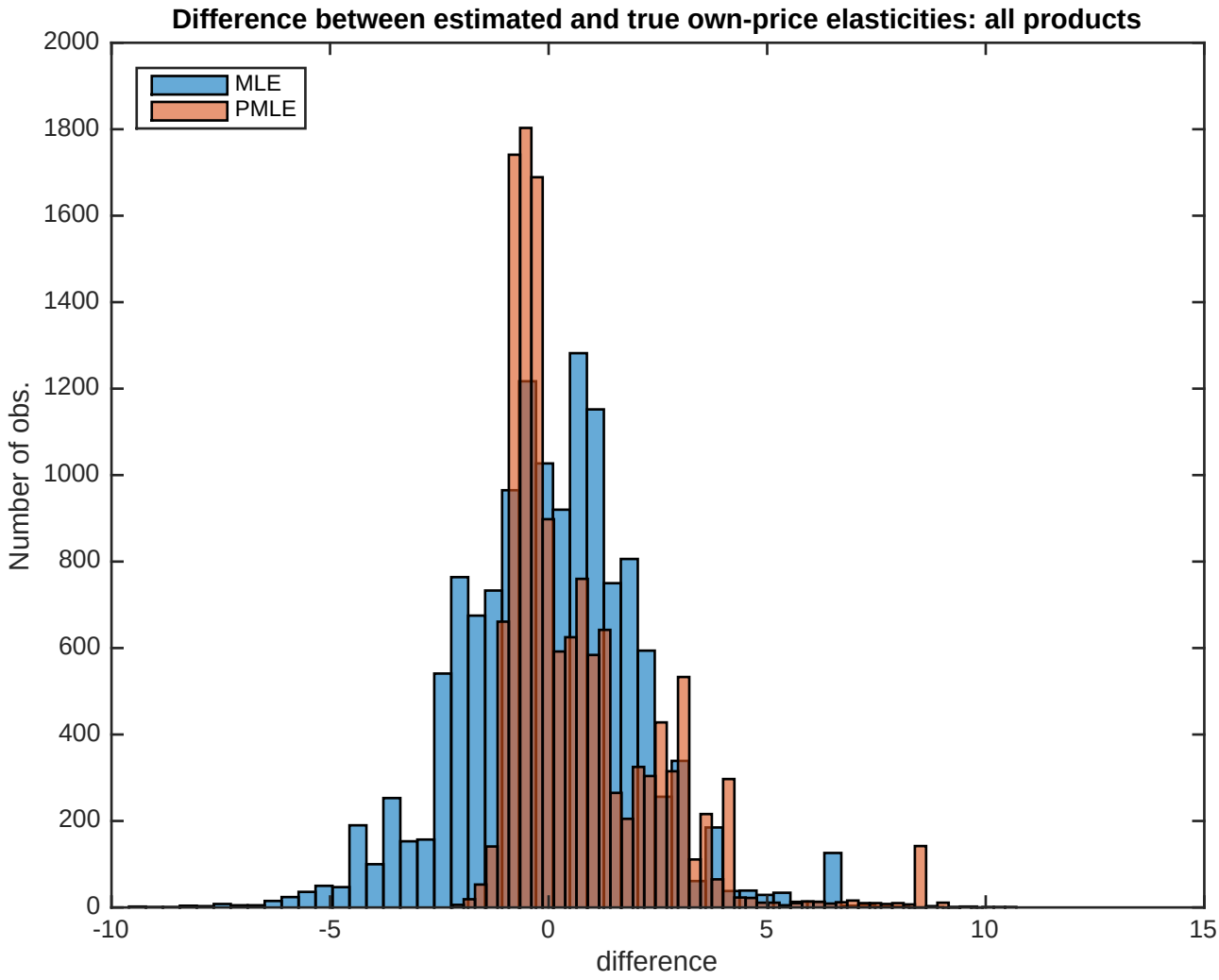

Difference between estimated and true price coefficient
MLE
PMLE
Number of obs.
140
120
100
80
60
40
20
0
-4
-3.5
-3
-2.5
-2
-1.5
-1
-0.5
0
0.5
difference

FIGURE 2


Difference between estimated and true own-price elasticities: all products
MLE
PMLE
Number of obs.
difference
2000
1800
1600
1400
1200
1000
800
600
400
200
0
-10
-5
0
5
10
15