Regularization Methods on Almost Ideal Demand System (AIDS)

Jinsoo Bae

The Ohio State University

July 9, 2018

Abstract

Almost Ideal Demand System (AIDS) has a flexible functional form that can first order approximate any sets of demand functions derived from utility maximizing behavior. This flexibility, however, comes with a cost of estimating numerously many parameters, hindering practical use of this model. In this paper, I suggest to use regularization methods on AIDS model to circumvent this curse of dimensionality. Monte Carlo simulation experiments, following Buse (1994), show that elastic net performs reasonably well under multicollinearity and sparsity environments.

1 Introduction

Almost Ideal Demand System (AIDS) (Deaton and Muellbauer, 1980) has a very flexible functional form that can first order approximate any sets of demand functions as its name suggests. However, this flexibility comes with a huge cost of estimating numerously many parameters, which makes the model impractical. For example, if the model estimates demand for J products, more than J^2 parameters have to be estimated to calculate own and cross-price elasticities of each product.

A traditional way to alleviate this problem is to put restrictions on parameters (Deaton and Muellbauer, 1980). Theoretical assumptions such as homogeneity and Slutsky symmetry are often imposed to reduce the number of parameters. Additional restrictions also can be put with any prior knowledge. For example, if we know that demand of a product does not depend of some other products, we can put zeros for cross-elasticity estimates.

An alternative way to alleviate the dimensionality issue is grouping products, often called multi-stage budgeting approach (Hausman et al., 1994; Hausman, 1996). In this model, similar products are grouped together in some sensible fashion, then AIDS model is applied for *each* group, dramatically reducing the number of parameters. A weakness of this approach, however, is that it severely restricts the substitution pattern of products without groups. For example, Hausman et al. (1994) segregated beer into three segments: premium, light, standard. In this case, price changes of a premium beer do not directly affect demand for light products. The effect only goes through group substitution; the price change of a premium product affects price index of premium segment, in turn, affects the demand for light group, and finally demand for light products. Thus, multi-stage budgeting may inaccurately estimate inter-group cross-price elasticities if the grouping is wrong.

Either using the traditional way or the alternative way, these two approaches are not free from critics that they put too strong prior assumptions. Theoretical restrictions such as homogeneity are often rejected in empirical studies (Deaton and Muellbauer, 1980). Using prior knowledge to put restrictions or grouping products may result in incorrect estimates if the prior knowledge is wrong. In this sense, it has been challenging to take advantage of the full flexibility of AIDS model without taking a risk of making wrong assumptions.

In this regard, I propose using regularization methods such as *lasso* or *elastic net* to estimate the AIDS model without making strong prior assumptions, but reducing the number of parameters to estimate. The only assumption needed to use these methods is *sparsity* which means some true parameters in the model are indeed zero[1]. Notice that sparsity assumption is realistic in AIDS model as demand of a product usually does not depend on every other product but only on several close products. Since lasso or elastic net have model selection ability to identify zero coefficients, under sparsity assumption, the two methods effectively reduce the number of parameters to estimate, alleviating the curse of dimensionality in AIDS model.

In this paper, I run Monte Carlo simulation experiments to investigate the relative performance of elastic net to OLS which is a usual estimator for the linearized AIDS model (LA/AIDS). In particular, I set three treatments where elastic net may have advantages over OLS. The first treatment was multicollinearity (MC) of prices, the second was sparsity (SP), and the last was both multicollinearity and sparsity (MCSP). The simulation results showed that elastic net estimator performed largely better than OLS estimator in all three treatments. The parameter choices in the simulations were close to Buse (1994).

The rest of the paper is organized as follows. In chapter 2, I briefly introduce regularization methods: lasso and elastic net. The AIDS model and the problem of "too many parameters" are discussed in chapter 3. Chapter 4 reports Monte Carlo Simulation results. Chapter 5 concludes the paper.

2 Regularization methods

2.1 Lasso

Lasso (least absolute shrinkage and selection operator) was proposed by Tibshirani (1996) to improve prediction accuracy and have model selection ability. Lasso minimizes the residual sum of squares subject to the sum of absolute values of the coefficients being less than

[1]Unlike exploiting prior knowledge, sparsity assumption need not specify which coefficients are zeros.

a constant. Because of the constraint in minimizing the residual sum of squares, Lasso produces some coefficients that are exactly 0, allowing model selection[2]. Although Lasso may increase bias, it more often than not improves overall prediction accuracy than OLS as it reduces the variance of the estimator, especially when the true model has many zero coefficients.

Definition 1 *Consider a linear regression model with n observation and p predictors $\mathbf{x_1}, ..., \mathbf{x_p}$. Let $\mathbf{y} = (y_1, ..., y_n)'$ be the response.*

$$\mathbf{y} = \beta_0 + \mathbf{x_1}\beta_1 + ... + \mathbf{x_p}\beta_p + \epsilon$$

let $X = (\mathbf{x_1}, ..., \mathbf{x_p})$ be the design matrix, where $\mathbf{x_j} = (x_{1j}, ..., x_{nj})'$. After location and scale transformation, we can assume that the response is centered and the predictors are standardized,

$$\sum_{i}^{n} y_i = 0, \quad \sum_{i}^{n} x_{ij} = 0 \quad and \quad \sum_{i=1}^{n} x_{ij}^2 = 1, \quad for \quad j = 1, 2, ..., p.$$

For any fixed non-negative λ the lasso estimator is defined by

$$\hat{\beta}(Lasso) = argmin_{\beta}\{L(\lambda, \beta)\} = |\mathbf{y} - X\beta|_2^2 + \lambda|\beta|_1$$

where

$$|\beta|_1 = \sum_{j=1}^{p} |\beta_j|$$

[2]Lasso has advantages over subset selection methods. Subset selection methods are unstable since it is a discrete progress; smaller changes in the data can result in very different model being selected. In contrast, lasso is a continuous shrinkage model that are less sensitive to changes in a data set.

2.2 Elastic net

Lasso has some weaknesses when the independent variables are highly correlated. If there is a group of variables that have high pairwise correlation, Lasso tends to select only one variable from the group and drop the others. As an attempt to overcome the weakness, Zou and Hastie (2005) proposed elastic net which has an additional L_2 penalty term in minimizing the residual sum of square. By adding this, elastic net tend to select all variables in an important group even though the variables are highly correlated.

2.2.1 Naive elastic net

Definition 2 *For any fixed non-negative λ_1, λ_2 the naive elastic net is defined by*

$$\hat{\beta}(naive\ elastic\ net) = argmin_{\beta}\{L(\lambda_1, \lambda_2, \beta)\} = |\mathbf{y} - X\beta|_2^2 + \lambda_2|\beta|_2^2 + \lambda_1|\beta|_1$$

where

$$|\beta|_2^2 = \sum_{j=1}^{p} \beta_j^2, \quad |\beta|_1 = \sum_{j=1}^{p} |\beta_j|$$

Although elastic net uses an additional L_2 norm in the penalization, it dose not impose higher estimation cost since the estimation can be reduced into Lasso. The residual sum of square and the L_2 norm can be combined and treated as a new residual sum of a new data set. Then Lasso technique can be applied to estimate elastic net.

2.2.2 Elastic net

Since naive elastic net uses additional L_2 penalty term, it shrinks the estimates to zero faster than Lasso. Hence, we need to correct this additional bias by inflating the estimates. This finally ends up with elastic net estimator.

Definition 3

$$\hat{\beta}(elastic\ net) = (1+\lambda_2)\hat{\beta}(naive\ elastic\ net)$$

3 Almost Ideal Demand System(AIDS)

3.1 Specification of AIDS

Almost Ideal Demand Estimation (AIDS) is a model for estimating demand for differentiated products. This model specifies a functional form which is flexible enough to (locally) approximate any cost function (i.e. expenditure function) derived from utility maximizing behavior. Moreover, the model uses a specific class of preferences that allows exact aggregation over individual consumer demand. This class of preferences known as PIGLOG can be represented by a cost function $c(u,p)$ for given utility u and price vector p,

$$logc(u,p) = (1-u)log\{a(p)\} + ulog\{b(p)\}a$$

In order for this functional form to be flexible enough to approximate any preferences, it must have enough parameters so that, at any single point, $\partial c/\partial p_i$, $\partial c/\partial u$, $\partial^2 c/\partial p_i \partial p_j$, $\partial^2 c/\partial u \partial p_i$, and $\partial^2 c/\partial u^2$ can be set equal to any arbitrary cost function. Deaton and Muellbauer (1980) take

$$\log\{a(p)\} = \alpha_0 + \sum_k \alpha_k logp_k + \frac{1}{2}\sum_k \sum_j \gamma^*_{kj} logp_k logp_j$$

$$log\{b(p)\} = log\{a(p)\} + \beta_0 \prod_k p_k^{\beta_k}$$

so that the AIDS cost function is written

$$logc(u,p) = \alpha_0 + \sum_k logp_k + \frac{1}{2}\sum_k \sum_j \gamma^*_{kj} logp_k logp_j + u\beta_0 \prod_k p_k^{\beta_k} \tag{1}$$

This cost function is homogeneous in p if the parameters satisfy $\sum_i \alpha = 1$, $\sum_j \gamma^*_{kj} = \sum_k \gamma^*_{kj}$ $= \sum_j \beta_j = 0$.

A merit of this specification is that it leads to a neat form of budget share function. By using Shephard' Lemma $\partial c(u,p)/\partial p_i = q_i$ and multiplying both sides by $p_i/c(u,p)$, we get

$$\frac{\partial logc(u,p)}{\partial logp_i} = \frac{p_i q_i}{c(u,p)} = w_i$$

where w_i is the budget share of good i. By differentiating (1), the budget share becomes

$$w_i = \alpha_i + \sum_j \gamma_{ij} logp_j + \beta_i u \beta_0 \prod p_k^{\beta_k} \tag{2}$$

$$\text{where } \gamma_{ij} = \frac{1}{2}(\gamma^*_{ij} + \gamma^*_{ji}) \tag{3}$$

Since the cost function $c(u,p)$, by definition, equals total expenditure x, the utility u can be recover by inverting this equality $c(u,p) = x$. The utility is $u = v(p,x)$, where $v(.,.)$ is the indirect utility function. Substituting u to (2) completes the AIDS demand functions in budget share form,

$$w_i = \alpha_i + \sum_j \gamma_{ij} logp_j + \beta_i log\{x/P\} \tag{4}$$

where P is a price index (often called exact price index) defined as

$$logP = \alpha_0 + \sum_k \alpha_k logp_k + \frac{1}{2}\sum_j \sum_k \gamma_{kj} logp_k logp_j \tag{5}$$

and the restrictions are now

$$\sum_{i=1}^{n} \alpha_i = 1 \ , \sum_{i=1}^{n} \gamma_{ij} = 0 \ , \sum_{i=1}^{n} \beta_i = 0 \tag{6}$$

$$\sum_{j} \gamma_{ij} = 0 \tag{7}$$

$$\gamma_{ij} = \gamma_{ji} \tag{8}$$

(6) means add-up constraints, (7) homogeneity , (8) symmetry.

3.2 Estimation of AIDS

The model can be estimated by substituting (5) to (4) and estimating the following non-linear budget share equation with error terms by MLE or other methods. All the parameters can be identified, provided that the number of observation large enough.

$$w_i = (\alpha_i - \beta_i \alpha_0) + \sum_j \gamma_{ij} logp_j + \beta_i \left(logx - \sum_k \alpha_k logp_k - \frac{1}{2} \sum_k \sum_j \gamma_{kj} logp_k logp_j \right) \tag{9}$$

Deaton and Muellbauer (1980) showed that the estimation can be done in an easier way if the prices are closely collinear. In this case, the exact price index P can be adequately approximated as proportional to some known index P^*, for example, Stone's index: $logP^* = \sum w_k logp_k$. By plugging in $P \approx \phi P^*$, the budget share equation (4) becomes

$$w_i = (\alpha_i - \beta_i log\phi) + \sum_j \gamma_{ij} logp_j + \beta_i log(x/P^*) \tag{10}$$

and can be estimated by a linear model such as OLS [3]. This linearized specification is often called linear approximated AIDS (LA/AIDS). Given a data set, $logp_j$ and $log(x/P^*)$ can be calculated from prices and incomes. The last step is to estimate ϕ. Let t denote each time

[3] When imposing cross-equation restrictions such as symmetry, SUR or 3SLS can be applied.

period when the budget share is observed. then we have

$$lnP_t = \alpha_0 + \sum_k \alpha_k logp_{kt} + \frac{1}{2}\sum_j \sum_k \gamma_{kj} logp_{kt} logp_{jt}$$

$$lnP_t^* = \sum_k w_{kt} logp_{kt}$$

As conventionally being done, it's possible to normalize all prices into 1(unity) at the base time period, say $t = 1$. Then $lnP_1 = \alpha_0$, $lnP_1^* = 0$. Thus, $lnP_1 - lnP_1^* = ln\phi = \alpha_0$.[4] Deaton and Muellbauer (1980) argued that α_0 can be pre-set with a priori value, since the parameter can be interpreted as the outlay required for a minimal standard of living.

3.3 Estimation issues in AIDS

As discussed above, AIDS is a flexible model that is grounded in a well-structured analytics and accommodates aggregations. So it has been a popular choice for many applied demand analyses. However, the advantages do not come without problems. First of all, the model has a large number of parameters in it, which are unlikely to be well determined with a usual size of data set. Specifically, if there are K number of products to estimate, then the number of parameters becomes $2K + K^2$ which grows very fast in K.

Deaton and Muellbauer (1980) suggest to put some retrictions to reduce the number of parameters. One obvious choice is to put theoretical restrictions such as homogenity and symmetry:(6), (7) and (8). These restrictions reduce the number of parameters into $2(K-1)+\frac{(k-1)+(k-2)}{2}$, but it still grows fast in K. Deaton and Muellbauer (1980) also suggest to put any restriction on γ_{ij} with prior knowledge. Since γ_{ij} $(i \neq j)$ has approximately the same sign as the compensated price elasticity, one can put $\gamma_{ij} = 0$ for independent products

[4] Of course, in reality, $ln\phi$ depends on the time period and cannot be a constant (Alston et al., 1994), causing inconsistency in linearized AIDS. However, this is not the main focus of my paper. I will focus on the performances of elastic net and OLS estimators under the same condition. For readers who are interested in the inconsistency problem, refer to Alston et al. (1994); Buse (1994); Pashardes (1993)

i and j.

Another way to reduce the number of parameters is grouping products in a sensible way and apply AIDS to each group (Multi-stage budgeting). By doing this, one can focus the substitution patterns only in each group, dramatically reducing the number of parameters. For example, Hausman et al. (1994) segregated beers into three categories (light mild, premium) and estimated AIDS model for each category. This method, however, severely restricts the substitution patterns of products in different groups. For example, price increase of a light beer does not directly affect the demand of a premium beer . The effect only goes through group substitution; the price increase of a light beer increases the price index of light beer, making the premium group more attractive and finally increases the demand for each product in the premium segment.

All these methods to reduce the number of parameters heavily depend on strong prior assumptions which may turn out to be incorrect. Deaton and Muellbauer (1980) mentioned that the theoretical restrictions such as homogeneity are often rejected in empirical studies. Other prior knowledge is not even contestable. Which products are independent? How can we ex-ante segregate product spaces into smaller groups in a sensible fashion before estimating the cross price elasticities?

In this regard, I suggest to use elastic net, a regularization method, to estimate AIDS model with less assumptions. First, I assume that the prices are close to collinear, which is a standard assumption for LA/AIDS. The second assumption is sparsity, which means that there are parameters whose true values are exactly zero. In AIDS model, sparsity can be interpreted that some products are independent of each other ($\gamma_{ij} = 0$ for some i and j). These two assumptions are weaker than traditional assumptions in that they do not ex-ante specify any substitution patterns.

I expect that elastic net have two advantages over OLS estimator. First, elastic net will be more robust to multicollinearity of prices as it includes L_2 penalty terms. Second, it will identify zero coefficients in the outcomes. These properties are well demonstrated in Zou and Hastie (2005), and the current paper is an application of elastic net to a demand estimation environment. In the next chapter, I run Monte Carlo simulation experiments to compare the performances of OLS and elastic net estimators.

4 Monte Carlo Simulation

In this chapter, I report Monte Carlo simulation results of OLS and elastic net for AIDS estimation. The simulation was run in three treatments. In the first treatment, the true model do not have zero coefficients, but the prices have multicollinearity problem (MC). In the second treatment, the prices are not highly correlated but sparsity (SP) is assumed with some parameters being zero. The third treatments assumes multicollinearity and sparsity (MCSP) at the same time.

4.1 The treatments of simulation

For the simulation, I closely followed the setting of Buse (1994) who did a thorough simulation on LA/AIDS model. In his setting, demands of 4 products are estimated by LA/AIDS model (10) and parameters are chosen consistent with consumer theories (homogeneity, symmetry). Following equations are reminders for share function and log expenditure.

$$w_i = \alpha_i + \sum_j \gamma_{ij} log p_j + \beta_i u \beta_0 \prod p_k^{\beta_k} \tag{11}$$

$$lnc(u, p) \equiv lnx = lnP + u\beta_0 \prod_k p_k^{\beta_k} \tag{12}$$

Buse (1994) set $u = 1$ and $\beta_0 = 1$ and unit prices, so that $ln(x/P) = 1$ at the base period. With this preset values we get $w_i = \alpha_i + \beta_i$, so only two of the variable can be freely defined. I set $w_i = (0.35, 0.2, 0.2, 0.25)$ and $\alpha_i = (0.15, 0.1, 0.3, 0.45)$ following Buse (1994). For γ_{ij}, only 6 free parameters can be chosen freely and others should be determined by homogeniety and symmetry restrictions. In the non-sparsity treatment, I set all of those parameters are non-zero while in the sparsity (SP) treatment, some of the parameters are set to be zero. The number of parameters are 24. For convenience, I will use a matrix form of parameters.

$$\begin{pmatrix} \alpha_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \beta_1 \\ \alpha_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \beta_2 \\ \alpha_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \beta_3 \\ \alpha_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \beta_4 \end{pmatrix}.$$

For generating regressors, I generated log prices through AR(1) process with a strong persistence, as often observed in the real data. I chose two types of correlations of the prices: high correlation(MC) and no correlation. I also generated log expenditure ($\log x$) from the AIDS expenditure function (12) by setting $\beta_0 = 1$ and letting u_t [5]

$$u_t = 1.0 + v_t \text{ where } v_t \sim N(0, 10^{-4}).$$

With all these parameters and regressors, I generated 100 observations (n) of each budget share(w_i) from the equation (11). By construction, the budget shares should sum up to 1. Finally, I added error terms $\epsilon_{it}$to each budget share, generating *observed* budget share $\tilde{w}_i$. This also should add up to 1, imposing $\sum_i \epsilon_{it} = 0$. Therefore, only 3 error terms need to be generated. I set the correlation among these 3 error terms $\rho_{12} = 0.5$, $\rho_{13} = 0.6$ and $\rho_{23} = 0.7$. [6] The variance of the error terms was set $2 * (10^{-6})$ to have reasonable variation in the budget shares. All these numbers and the data generating process are very similar to Buse (1994) except I have sparsity cases.

[5]Buse (1994) mentions that this specification introduce variation in utility that might be realistically expected in a time series context. On the contrary, in Alston et al. (1994)'s study, u was fixed at 0.5

[6]This correlation actually is not important for estimating each equation separately. However, these take a role when someone wants to estimate the 4 equations at the same time with SUR or 3SLS

4.2 Multicollinearity(MC)

In this treatment the prices are collinear and all the parameters are non-zero. The correlation of log prices are set as $r_{12} = 0.99$, $r_{13} = r_{23} = 0.98$, $r_{14} = r_{24} = r_{34} = 0.97$, following Buse (1994). This magnitude of correlation in prices are common, and Stone's price index P^* can be a good approximation of exact price index P with this high collinearity. The following is the true parameters in this treatment.

$$\begin{pmatrix} \alpha_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \beta_1 \\ \alpha_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \beta_2 \\ \alpha_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \beta_3 \\ \alpha_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \beta_4 \end{pmatrix} = \begin{pmatrix} 0.15 & 0.0875 & -0.0300 & -0.0100 & -0.0475 & 0.2 \\ 0.1 & -0.0300 & 0.0700 & -0.0200 & -0.0200 & 0.1 \\ 0.3 & -0.0100 & -0.0200 & 0.0500 & -0.0200 & -0.1 \\ 0.45 & -0.0475 & -0.0200 & -0.0200 & 0.0875 & -0.2 \end{pmatrix}$$

In this specification, I expect that OSL will suffer from Multicollinearity but elastic net would be more stable than OLS since it has the property of the Ridge equation. The following graph shows price changes in the data generation. The prices are highly correlated and persistent. Since prices are following AR(1) process, it tends to revert to zero. In this specific realization, the log prices move around between 0.10 $\sim$ -0.30, which is quite reasonable changes consistent with most price data sets[7].

[7] For example, the prices of agricultural products or animal products which demand is often studied wtih AIDS model have similar price changes as the simulation.

Figure 1: Prices

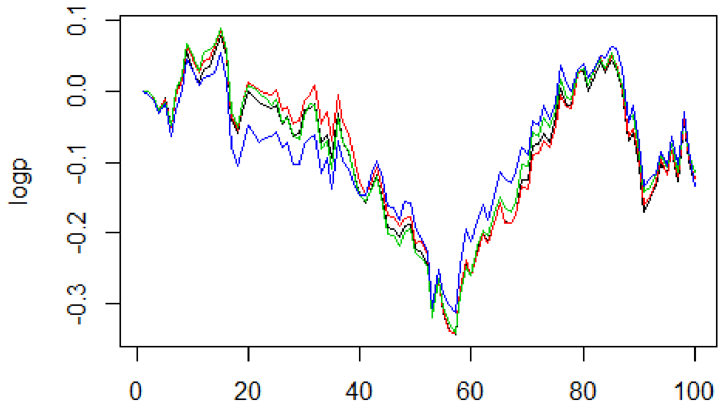
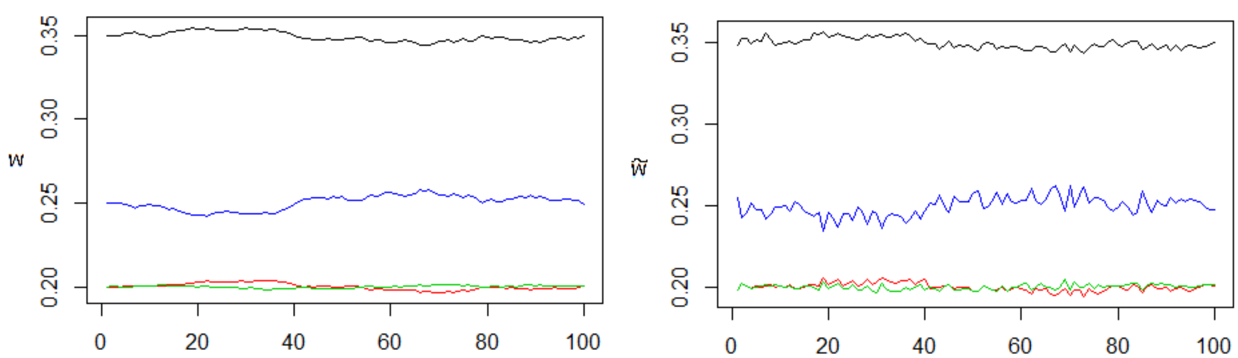


Figure2 (left) shows the series of budget share. Since the prices are moving in the same directions, the substitution effect is minimal, and the budget share does not change volatilely. Figure 2 (right) shows the observed budget share after error term is added. The maximum deviation of observed share form the true share did not exceed 2% and in average the deviation in general was very small.

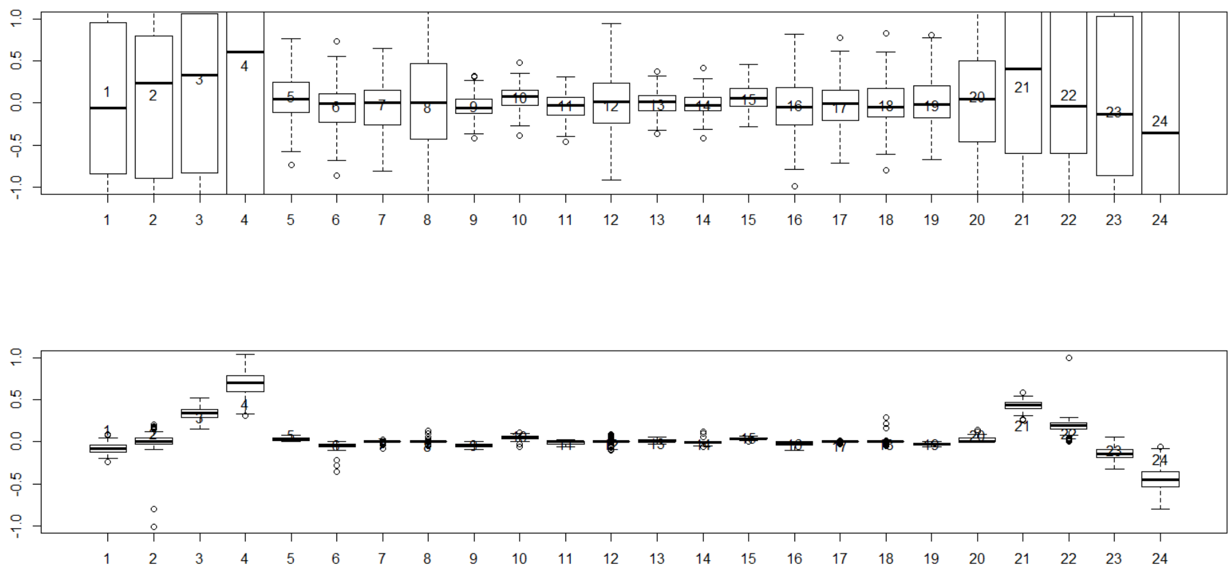
Figure 2: Budget Share



With the generated regressors and error terms, I estimated the parameters with OLS and Elastic net 100 times each. For elastic net, I choose small portion of L_2 norm as $\frac{\lambda_1}{\lambda_2+\lambda_1} = 0.9$, and the tuning parameters were set to have least MSE by 10-fold cross validation. Figure 3 shows the box-plots of estimated parameters from OLS and elastic net. It's apparent that the OLS estimators have much bigger variance due to multicollinearity of the regressors.

Specifically, this model has two sources of Multicollinearity. First, as I mentioned, the prices are highly collinear. On top of this, I found that the constant term and the $log\{x/P\}$ terms are collinear since the income and the price index does not change dramatically over time. Elastic net, however, is quite robust to multicollinearity as it has the property of Ridge regression. Even though it is a biased estimators, the bias seems more bearable than the huge variance of OLS estimators in this treatment.

Figure 3: Box-plots of estimates(MC)



The box plots of OLS and elastic net estimators. The locations of the numbers inside the plot are the values of true parameters. The order (1-24) follows $\alpha.$, $\gamma_{.1}, \gamma_{.2}, \gamma_{.3}, \gamma_{.4}, \beta$.

Of course, this result does not mean that elastic net always performs better than OLS. Basically OLS is an unbiased estimator, thus it gets the true parameters right in average. Therefore, when having enough number of observations to reduce the variance significantly, there is no reason to avoid OLS. In addition, we know that elastic net is a biased estimator, so it is easy to find some environments where OLS can do better than elastic net. I could reduce the variance of OLS estimators by either increasing the variation of the regressors or reducing the variance of error terms. However, in order for OLS estimator to have significantly better outcome, it requires unrealistic modification in the data. For example, I needed to increase variation of the regressors (price or income) more than 10 times, which leads the budget shares of each products to fluctuate severely. When I tried to reduce the variance of error terms, it required me to reduce the error term more than 10 times smaller to the point that the error term is almost negligible.

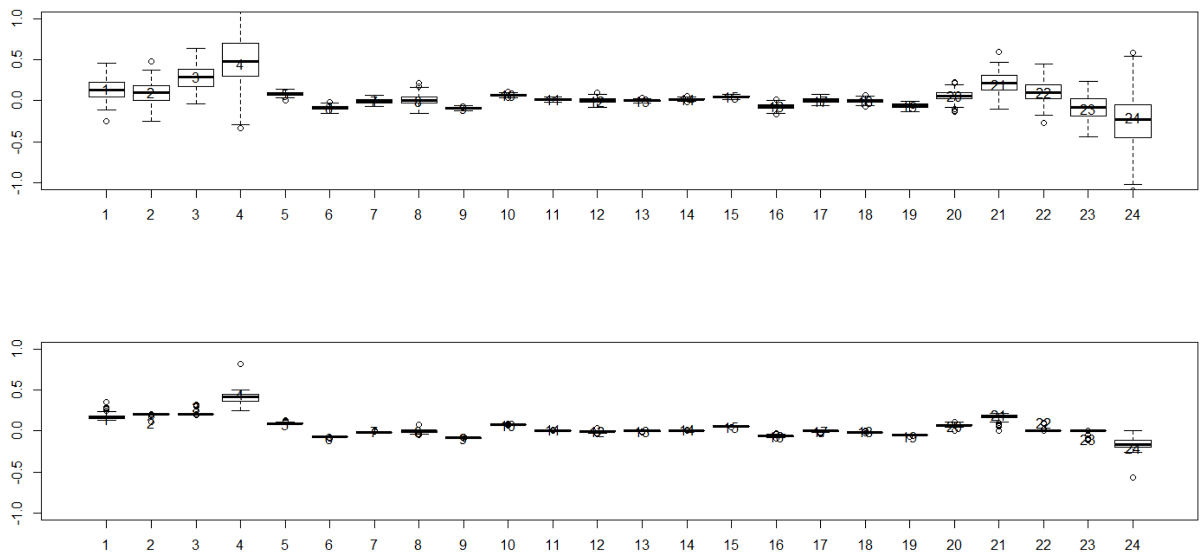
4.3 Sparsity(SP)

In this treatment, the prices are no longer collinear but some parameters are truly zero. I generate completely independent AR(1) prices which is very unrealistic. In addition I set ten times bigger variation in the utility ($u \sim N(0, 10^{-3})$) so that the multicollinearity of constant and the expenditure term can be alleviated. Finally, I set some parameters whose values are exact zero. Specifically, I set $\gamma_{13} = \gamma_{14} = \gamma_{23} = 0$ which means these are (approximately) independent goods. Thus, the parameters are

$$\begin{pmatrix} \alpha_1 & \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \beta_1 \\ \alpha_2 & \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} & \beta_2 \\ \alpha_3 & \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \beta_3 \\ \alpha_4 & \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} & \beta_4 \end{pmatrix} = \begin{pmatrix} 0.15 & 0.0875 & -0.0875 & 0 & 0 & 0.2 \\ 0.1 & -0.0875 & 0.0700 & 0.0175 & 0 & 0.1 \\ 0.3 & 0 & 0.0175 & 0.0500 & -0.0675 & -0.1 \\ 0.45 & 0 & 0 & -0.0675 & 0.0675 & -0.2 \end{pmatrix}$$

In this treatment, I expect that OLS will do better than the previous treatment since SP treatment does not have severe multicollinearity. However, it is clear that OLS will never choose any zero estimate. In contrast, I expect that the elastic net will choose several zero estimates. The figure 4 shows the box-plots of the estimates of the two estimators. As expected, the OLS has much less variance compared to MC treatment. However, it still have higher variances than elastic net estimators. One obvious reason for this higher variance is that OLS overfits the model by estimating true-zero parameters. In this treatment, many of the true parameters are zero, so model selection helps to reduce the variance of the estimators. Indeed, elastic net has much less variance than OLS as it drops some regressors in estimation.

Figure 4: Box-plots of estimates(SP)



Elastic net, however, does not always choose the true model. For example, it selects β_2, β_3 (22, 23) to be zero many times although they are non-zero coefficients. I counted the frequency of choosing the zero parameters in following matrix. The left matrix is true values of coefficients and the right matrix means the frequency that elastic net choose zeros out of 100 simulation. It is not surprising that elastic net does not always select the true zero value. First of all, elastic net is not selection consistent (Zou and Zhang, 2009). Second, the number of observations was only 100 (for each equation). With finite sample size, model selection methods can fail to identify the true model.

$$\begin{pmatrix} 0.15 & 0.0875 & -0.0875 & \mathbf{0} & \mathbf{0} & 0.2 \\ 0.1 & -0.0875 & 0.0700 & 0.0175 & \mathbf{0} & 0.1 \\ 0.3 & \mathbf{0} & 0.0175 & 0.0500 & -0.0675 & -0.1 \\ 0.45 & \mathbf{0} & \mathbf{0} & -0.0675 & 0.0675 & -0.2 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 0 & 0 & \mathbf{43} & \mathbf{44} & 1 \\ 0 & 0 & 0 & 1 & \mathbf{2} & 95 \\ 0 & \mathbf{3} & 1 & 0 & 0 & 92 \\ 0 & \mathbf{54} & \mathbf{31} & 0 & 1 & 13 \end{pmatrix}$$

4.4 Multicollinearity and Sparsity (MCSP)

The last treatment includes both multicollinearity and sparsity. Presumably, this is the most common situation in the real world data, since the price are often highly collinear, and the demand of a product is usually affected by only a few close products. In this treatment, I used the correlation of the MC treatment and the same parameters in the SP treatment. Theoretically, this treatment is the most favaorable to elastic net relative to OLS.

Figure 5: Box-plots of estimates(MCSP)

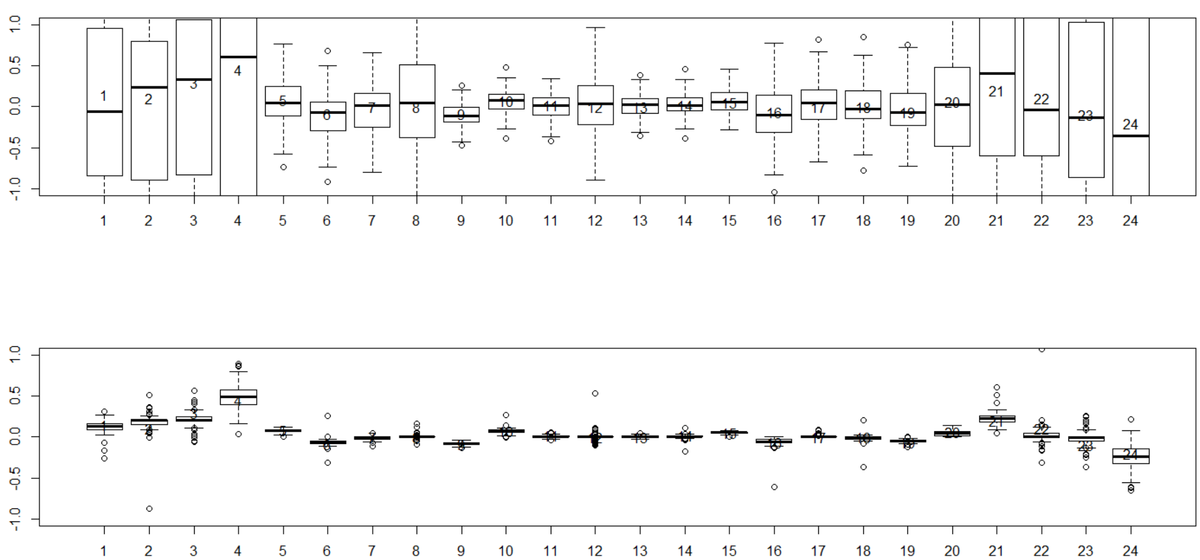


Figure 5 is the box plots of the coefficients of the two estimators. First of all, we can see that OLS estimators has higher variances as expected. Since sparsity of the coefficient does not affect OLS estimator, the box plot of OLS is almost the same with MC treatment. The elastic net has similar variances as in MC, but its bias is much small than MC. This is because there are many true zero values, the elastic net easily choose the zeros with smaller penalty terms. I counted the number of elastic net choosing zero coefficients. Although elastic net is not selection consistent, we can see that the elastic net tends to choose the true zeros quite frequently in MCSP treatment.

$$\begin{pmatrix} 0.15 & 0.0875 & -0.0875 & \mathbf{0} & \mathbf{0} & 0.2 \\ 0.1 & -0.0875 & 0.0700 & 0.0175 & \mathbf{0} & 0.1 \\ 0.3 & \mathbf{0} & 0.0175 & 0.0500 & -0.0675 & -0.1 \\ 0.45 & \mathbf{0} & \mathbf{0} & -0.0675 & 0.0675 & -0.2 \end{pmatrix} \rightarrow \begin{pmatrix} 0 & 1 & 0 & \mathbf{21} & \mathbf{60} & 0 \\ 0 & 3 & 1 & 17 & \mathbf{20} & 25 \\ 0 & \mathbf{42} & 40 & 1 & 2 & 16 \\ 0 & \mathbf{77} & \mathbf{69} & 16 & 19 & 2 \end{pmatrix}$$

Overall, elastic net shows better performance than OLS estimators in the environment of multicollinearity and sparsity.

5 Conclusion

In this paper, I compare the estimation performance of OLS and elastic net for LA/AIDS model in several treatments. The Monte Carlo simulation experiments show that elastic net do a better job than OSL under multicollinearity and sparsity environment.

One might argue that LA/AIDS is not as accurate as the full model of AIDS, thus, elastic net is not useful in estimating the AIDS model. However, LA/AIDS model is still popular due to its tractability. Moreover, elastic net is not restricted to a linear model but it can be also applied to MLE type of estimators by adding L_1 and L_2 penalty terms. Future research may investigate the relative performance of elastic net or other regularization methods on the full model of AIDS estimation.

References

Alston, J. M., K. A. Foster, and R. D. Gree (1994). Estimating elasticities with the linear approximate almost ideal demand system: some monte carlo results. *The review of Economics and Statistics*, 351–356.

Buse, A. (1994). Evaluating the linearized almost ideal demand system. *American journal of agricultural economics 76*(4), 781–793.

Deaton, A. and J. Muellbauer (1980). An almost ideal demand system. *The American economic review 70*(3), 312–326.

Hausman, J., G. Leonard, and J. D. Zona (1994). Competitive analysis with differenciated products. *Annales d'Economie et de Statistique*, 159–180.

Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In *The economics of new goods*, pp. 207–248. University of Chicago Press.

Pashardes, P. (1993). Bias in estimating the almost ideal demand system with the stone index approximation. *The Economic Journal 103*(419), 908–915.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

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*(2), 301–320.

Zou, H. and H. H. Zhang (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics 37*(4), 1733.