Topics in Econometrics and Statistics

Monte Carlo Analysis of Fixed Coefficients BLP Model with High-Dimensional Product Characteristics and Instruments

Prepared by: Nebi Simsek
Student ID: 3395869
Instructor: Prof. Dr. Joachim Freyberger
Date: January 9, 2023

Abstract

Gillen, Shum, and Moon (2014) offer a two-stage approach when solving high dimensionality issues in random coefficient BLP models. In the first stage, they collect mean utility levels with penalized GMM. In the second stage, they select the relevant characteristics and instruments using the consistent mean utilities collected in the first stage. Finally, they obtain a uniformly consistent result for the price coefficient using selected instruments and characteristics. One of the special and easier cases of the BLP model is assuming the coefficients are not random. Keeping the coefficients fixed is an unpopular practice since it has undesirable elasticity properties. However, it is still a good starting point to understand the behavior of random coefficient models when there is high dimensionality. Moreover, it can still be used in more relative settings. In this report, I considered this special case of the BLP model and analyze the performance of Gillen et al. (2014) procedure with Monte Carlo Simulations. My findings are similar to the results of Gillen et al. (2014) with lower bias and variance as expected from logit model. However, I observed weaker performance when choosing relevant characteristics. My initial results do not clarify the importance of selecting instruments. As an additional analysis to Gillen et al. (2014), I increased the number of instruments, I have checked the performance of the model with weak instruments and I have increased the intensity of endogeneity in some simulations. Selecting instruments became much more important as I increased the number of instruments and intensity of endogeneity. Unfortunately, the improvement became ambiguous with weak instruments. Overall, the procedure Gillen et al. (2014) follows outperforms alternatives I have tested when there are many characteristics and many instruments.

Contents

1 Introduction

Modeling demand is not a straightforward process. There is a simultaneity issue between demand and price which needs more complicated approaches. Berry (1994) and Berry et al. (1995) came up with such an approach called the BLP model. The model uses the random utility structure as its basis. After some assumptions and calculations, they conclude a model which takes market-level information as input and produces meaningful price coefficients. Moreover, since the model uses random coefficients, it has good elasticity properties, unlike the regular logit model.

In such a model that takes market-level data as input, we may observe a limited number of products and a limited number of markets. On the other hand, we may have many characteristics, but having more characteristics than observation cause unidentified parameters. Moreover, if the product characteristics do not differ in different markets, the number of characteristics that allow identified results is bounded by the number of unique products. Luckily, If we think a limited number of characteristics are relevant, we can deal with this high dimensionality issue.

To solve this issue Gillen et al. (2014) offer a two-stage approach. In the first stage, they collect mean utilities and endogenous parts consistently with penalized GMM. The second stage consists of three steps. In the first step they use the collected mean utilities and select the relevant characteristics. Following the procedure of Belloni et al. (2013), as a second step they select those characteristics correlated with price and get the union of the characteristics selected in these first two steps. In the third step, they select relevant instruments with a Lasso penalty. Finally, using these selected instruments and characteristics they obtain a uniformly consistent result for the price coefficient with an unpenalized GMM.

One of the special and easier cases of the BLP model is assuming the coefficients are not random. Such a model is unpopular since it has undesirable elasticity properties. However, it is still a good starting point to understand the behavior of random coefficient models when there is high dimensionality. Moreover, it can be used in more relative settings such as a product with negligible differences in its characteristics and price over different brands. In this report, I considered this special case of the BLP model and analyze the performance of Gillen et al. (2014) procedure with Monte Carlo Simulations.

Using the logit model to analyze the performance of Gillen et al. (2014) made the first stage redundant. We can directly collect mean utility levels by subtracting the logarithm of the outside option's market shares from the logarithm of the relative product's market share. The rest of the steps followed exactly as described by Gillen et al. (2014) and verified the performance of the procedure. In addition to their analysis, I have also controlled the performance of the procedure for the number of instruments, weak instruments, and more intense endogeneity. From Belloni et al. (2013) we know that the initial two steps lead us to reach uniformly consistent results. On the other hand, the need for the third step that we select instruments became more explicit as we increased the number of instruments. Unfortunately, this improvement is lost when I use weak instruments.

In the next part, I will talk about the structural model more explicitly and describe the Triple-Lasso procedure. In the third section, I will mention the data generating process I used and talk about the result of Monte Carlo Simulations. Finally, in a brief conclusion, I will mention what was interesting for me and what kind of future topic this project can lead me to.

2 Structural Model and Triple-Lasso Procedure

2.1 Structural Model

To build the structural model, first, we describe the utility of the person i. In each market $m = 1, \ldots, M$ person i chooses among $J + 1$ alternatives that also possess an outside good, $j = 0$. So, the utility of the person i choosing product $j = 1, \ldots, J$ in market m will be equal to

$$u_{ijm} = x'_{jm}\beta_x + p_{jm}\beta_p + \xi_{jm} + \epsilon_{ijm} \tag{1}$$

Here x'_{jm} is a vector of K characteristics, p_{jm} is price, and ξ_{jm} is the endogenous part correlated with the price. β_x is a vector of coefficients for characteristics and β_p is the price coefficient. ϵ_{ijm} stands for Type-I Extreme Value distributed i.i.d. errors. Thanks to this assumption about the errors, we can write the probability of the agent i buying the product j in market m as

$$P(y_{ijm} = 1 | X_m, p_m, \xi_m) = \frac{exp(x'_{jm}\beta_x + p_{jm}\beta_p + \xi_{jm})}{1 + \sum_{k=1}^{J} exp(x'_{km}\beta_x + p_{km}\beta_p + \xi_{km})} \tag{2}$$

Unlike the random coefficient model, this probability equation doesn't depend on individual i. So the aggregation of probability to get the share of product j in market m will give the same equation. Which can be written as

$$s_{jm} = \frac{exp(x'_{jm}\beta_x + p_{jm}\beta_p + \xi_{jm})}{1 + \sum_{k=1}^{J} exp(x'_{km}\beta_x + p_{jm}\beta_p + \xi_{jm})} \tag{3}$$

Moreover, Berry (1994) showed that there is one to one relationship between shares and mean utility. So, knowing the market level shares, we can obtain δ_{jm} with an inverse function. Unlike the random coefficient model, for the logit model to obtain δ_{jm} involves simple calculation instead of a contraction. Again Berry (1994) mentions this as a special case and shows that we can obtain the mean utility level by subtracting the logarithm of the share of outside good from the logarithm of the share of good j.

$$\log s_{jm} - \log s_{j0} = \delta_{jm} = x'_{jm}\beta_x + p_{jm}\beta_p + \xi_{jm} \tag{4}$$

Where s_{j0} stands for the outside option. Obtaining δ_{jm}, now we have a clear model where we possess all the relevant data. However, there is also an unobserved endogenous part that we have to deal with. To solve endogeneity, we can use instrumental variables which are correlated with the price. So, we need to introduce z_{jm} as the set of instruments with L dimensions where $(L+K) \geq (K+1)$ since there is a price variable in addition to K number of characteristics.[1] Then we can apply two-stage least squares. In the first stage, we obtain predicted X matrix with the following projection.

$$\hat{X} = Z(Z'Z)^{-1}Z'X$$

Here, X is $N \times (K+1)$ and Z is $N \times (L+K)$ matrix where N equals to $J \times M$ amount of observation. In the second stage, we can get the relative coefficients by regressing the mean utility vector on the estimated X matrix.

$$\hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'\delta$$

2.2 Selection with Triple-Lasso

In my analysis, I assume product characteristics stays the same over different markets. With this assumption, to have identified parameters the number of products J should at least be equal to $K+1$ amount of characteristics. If we have more characteristics than the amount of the unique products and we have sparsity, we may drop the irrelevant characteristics to have an identified case.

To determine this irrelevant characteristics and drop them in high-dimensional cases Gillen et al. (2014) uses a two-stage procedure. In the first stage, they use all the variables to obtain δ. In logit model, this stage does not require the use of characteristics. We only need to know the share of the products in the relative market. Subtracting the sum of all shares in the market from one, we can obtain the share of the outside good. Finally, δ values can be collected as described in 4.

In the second stage I will follow the three-step procedure described by Gillen et al. (2014). In the first step δ will be regressed on characteristics with Lasso penalty. As a result of this step, we will select the characteristics that drive the mean utilities.

[1] Here, I kept L number of instruments as a separate entity from the K number of characteristic. In following pages I will control for instruments by increasing the L. Keeping instruments separated from characteristics will prevent the confusion.

$$\hat{\gamma}_x = \underset{\gamma_x}{argmin} = \sum_{m=1}^{M}\sum_{j=1}^{J}(\delta_{jm} - x'_{jm}\gamma_x)^2 + \lambda_x||\gamma_x||_1$$

To follow Double Lasso Procedure by Belloni et al. (2013) we will regress price variable on characteristics again with the Lasso Penalty.

$$\hat{\gamma}_p = \underset{\gamma_p}{argmin} = \sum_{m=1}^{M}\sum_{j=1}^{J}(p_{jm} - x'_{jm}\gamma_p)^2 + \lambda_p||\gamma_p||_1$$

With this second step we will collect the characteristics correlated with the price. Say X_1 is the characteristics selected in the first stage, X_2 is the characteristics select in the second stage, The union of these two sets of characteristics, X^*, will be the set of characteristics that we will use for post estimation.

As a third and final step we will select the relevant instruments. To select relevant characteristics, now we regress the price on the set of instruments with the lasso penalty.

$$\hat{\zeta}_p = \underset{\zeta_p}{argmin}\sum_{m=1}^{M}\sum_{j=1}^{J}(p_{jm} - z'_{jm}\zeta_p)^2 + \lambda_z||\zeta_p||_1$$

Say the selected relevant instruments with this step is Z_1. Since we only assume the price is endogenous, taking the union of X^* and Z_1 will be the set of instruments, Z^*, that we will use for the post estimation.

After collecting X^* and Z^*, now we can implement post-selection stage. We will follow $2SLS$ procedure as described above with the selected sets of characteristics. In the first stage again we obtain the estimated X^*.

$$\tilde{X}^* = Z^*(Z^{*\prime}Z^*)^{-1}Z^{*\prime}X^*$$

And in the second stage we obtain the coefficients.

$$\tilde{\beta}_{2SLS} = (\tilde{X}^{*\prime}\tilde{X}^*)^{-1}\tilde{X}^{*\prime}\delta$$

3 Monte Carlo Simulations

3.1 Data Generating Process

For the data generating process, I used slightly different structures than Gillen et al. (2014). I used 20 products in 5 different markets. Each product exists in each market but product characteristics do not differ over different markets even though prices may differ. This approach is used since it is closer to what we usually observe in real life. A better approach would be to use closely correlated prices and characteristics for the same products across different markets, but for simplicity, I used an easier setting as described above. For the analysis, I used a different number of characteristics, $K \in \{10, 19, 50, 100\}$, to measure the performance of the model. I allow a certain level of correlation among characteristics as followed by Gillen et al. (2014).

$$X_{jm} \sim N(0, \Sigma_X); \quad \Sigma_{X,\{i,j\}} = 0.8^{|i-j|}/8 \tag{5}$$

where i and j stand for respective rows and columns for the variance-covariance matrix of characteristics. Similarly for instruments;

$$Z_{jm} \sim N(0, \Sigma_Z); \quad \Sigma_{Z,\{i,j\}} = c_1^{|i-j|} \quad \text{s.t.} \quad i, j \leq r \tag{6}$$

Here, the first element of the variance-covariance matrix and data generated with this part belongs to price. c_1 determines the correlation between instruments and the price variable. By setting a smaller c_1 we may get weaker instruments. Finally, $r-1$ stands for the number of relevant instruments. For irrelevant instruments, a mean of zero and a variance of 1 is used. Now we can generate the price with the following.

$$\xi_{jm} \sim N(0, c_2); \quad \nu_{jm} \sim U(0, 1) \tag{7}$$

$$p_{jt} = \xi_{jt} + Z_{jt,\{,1\}} + [1, 1, 0, 0, \dots, 0, 0]X_{jt} + \nu_{jt} \tag{8}$$

Where c_2 represent the intensity of endogeneity since $cov(p_{jm}, \xi_{jm}) = var(\xi_{jm})$. By setting a higher c_2 we can observe how the estimator behaves for more intense endogeneity. With that, we have all the data to characterize the utility of agent i.

$$u_{ijm} = [4, 4, 4, 4, 0, \dots, 0]X_{jm} + (-5) \times p_{jm} + \xi_{jm} + \epsilon_{ijm}; \quad \epsilon_{ijm} \sim \text{Type I Extreme Value} \tag{9}$$

3.2 Compared Estimators

As followed in Gillen et al. (2014), I will compare a couple of different estimators to understand the performance of the Triple-Lasso procedure. The benchmark is represented with the best possible selection. This estimator is called 'Oracle Benchmark' and uses all the relevant characteristics and all the relevant instruments as if it is known a priori while all the irrelevant characteristics and instruments are excluded.

Another alternative estimator uses all the characteristics and instruments generated during DGP. This estimator is called 'Unpenalized 2SLS' as it doesn't include any step to decrease the number of characteristics or instruments. This estimator is only defined when we have a smaller or equal number of characteristics than the number of unique products. Since K represents the total number of characteristics, $K = 19$ is the bound for the defined 'Unpenalized 2SLS' estimator in our simulations.

I add another two estimators that don't follow the Triple-Lasso procedure exactly to show that all the steps are required. The first one is Post-Single Lasso which only selects relevant control variables and omits the second and third steps. Different than Gillen et al. (2014), I employed Post-Double Lasso instead of Penalized GMM since the first stage became redundant. Post-Double Lasso estimator follows only the Double Lasso procedure offered by Belloni et al. (2013) and omits the selection of instruments.

3.3 Results

The estimators perform similarly to results in Gillen et al. (2014). For a lower number of characteristics, all estimators except Post-Single Lasso perform well. Post-Single Lasso always omits some relevant characteristics so it produces weak results. For a higher number of characteristics, Post-Triple Lasso preserves its good predictions. On the other hand, the Post-Double Lasso estimator which omits the Instrument Selection step also performs similarly to the Post-Triple Lasso estimator. This initial result may raise questions about the need for the third step. However, we will see the importance of this step as we increase the number of instruments in subsequent analyses. In Table 1, I used 10 instruments and out of these 10 instruments only 5 of them were relevant. The Data Generated as described in 3.1. I set $c_1 = 0.7$ and $c_2 = 0.5$.

Although the predictions for price coefficient are better in my analysis, the selection of relevant characteristics in Table 2 performs poorly compared to Gillen et al. (2014). This is probably related to the choice of penalty term. Gillen et al. (2014) briefly mentions that they use block cross-validation technique. They may also be controlling for a wider range of λ values. For simplicity, I used 10-fold cross-validation in my analysis and for feasibility,

Table 1: Sampling Properties of Estimated Price Parameters

	K = 10	K = 19	K = 50	K = 100
	Bias for β_p			
Oracle Benchmark	0.015	0.019	0.017	0.016
Unpenalized 2SLS	0.041	0.048	-	-
Post-Triple Lasso	0.012	0.011	0.004	0.007
Post-Single Lasso	0.236	0.266	0.288	0.317
Post-Double Lasso	0.037	0.040	0.034	0.038
	Standard error for β_p			
Oracle Benchmark	0.073	0.071	0.074	0.072
Unpenalized 2SLS	0.069	0.070	-	-
Post-Triple Lasso	0.097	0.107	0.120	0.120
Post-Single Lasso	0.306	0.322	0.348	0.386
Post-Double Lasso	0.088	0.096	0.109	0.111
	Root Mean Squared Error for β_p			
Oracle	0.071	0.071	0.077	0.071
Unpenalized	0.077	0.084	-	-
Post-Triple	0.100	0.110	0.118	0.118
Post-Single	0.386	0.418	0.452	0.500
Post-Double	0.095	0.105	0.114	0.118

The table compares the performance of 5 different estimators for a different number of characteristics. Out of the characteristics on each column, only 4 of them is relevant. To measure the performance; Bias, Standard Error, and Root Mean Squared Error for β_p are used. The Data Generating Process Described in the 3.1. I used 5 relevant and 5 additional irrelevant instruments for each cases. Here I set $c_1 = 0.7$ and $c_2 = 0.5$ which represents the relevance of the most relevant instrument and intensity of endogeneity respectively. The results is computed from 2500 simulations where I used 20 unique products and 5 different markets. All the products are included in all the markets making 100 observations. Oracle Benchmark uses only the relevant characteristics and instruments. Unpenalized 2SLS uses all characteristics and instruments without any selection. Post-Triple Lasso is the estimator described in this report. Post-Single Lasso uses only the first, and Post-Double Lasso uses the first two steps of the described procedure.

I keep the range for λ limited. So, improving the performance of the penalty term can be possible in a more extensive analysis

For $K \in \{50, 100\}$, the total number of selected characteristics exceeds 19 in some simulations. But the numbers were negligible (29 and 71 out of 2500 respectively). When reporting sampling properties I omit those cases, but I include them in Table 2. I used the penalty term that produces the lowest cross validated error out of all alternatives. For an appropriate analysis, I keep this choice fixed. However, it may be a good idea to further adjust the penalty term if we have too many characteristics and we think we have enough sparsity.

Table 2: Simulation Results on Characteristics Selection

	Frequency of Selecting Important Attributes			
A	K = 10	K = 19	K = 50	K = 100
Oracle Benchmark	100%	100%	100%	100%
Unpenalized 2SLS	100%	100%	-	-
Post-Triple Lasso	89%	84%	78%	75%
Post-Single Lasso	45%	39%	32%	29%
Post-Double Lasso	89%	84%	78%	75%
B	Average number of null attributes Selected			
Oracle Benchmark	0.00	0.00	0.00	0.00
Unpenalized 2SLS	6.00	15.00	-	-
Post-Triple Lasso	2.13	3.50	5.25	6.39
Post-Single Lasso	1.48	2.39	3.21	3.63
Post-Double Lasso	2.13	3.50	5.25	6.39

The table summarizes the simulation results of Characteristics Selection. Part A summarizes the percentage of selecting all relevant attributes with respective procedure. If all 4 of the relevant attributes are selected in each simulations,the result is 100%, if all relevant attributes are selected in half of the simulations, it is 50 % etc. Part B summarizes the average number of selected null attributes. The total number of null attributes is $6, 15, 46, 96$ respectively. The number in the table indicates average over 2500 simulations. Oracle Benchmark is the infeasible estimator that uses only the relevant attributes and omit the irrelevant ones. Unpenalized 2SLS uses all the characteristics generated during DGP. Post-Triple Lasso and Post-Double Lasso, both uses first and second selection step, so the results are equal. Post-Single Lasso only includes only the first selection step.

The frequency of selection of relevant instrumental variables is similar to the selection of characteristics for the Post-Single Lasso estimator. For this initial analysis, I used 10 instruments in which 5 of them were relevant. On 38.5% of the cases, all the relevant instruments are included in the model, and on average 1.3 irrelevant instruments are added. Table 3 reports, the relevance of individual instruments and how often we include them in the model with the choice of penalty term.

Table 3: Simulation Results on Instrument Selection

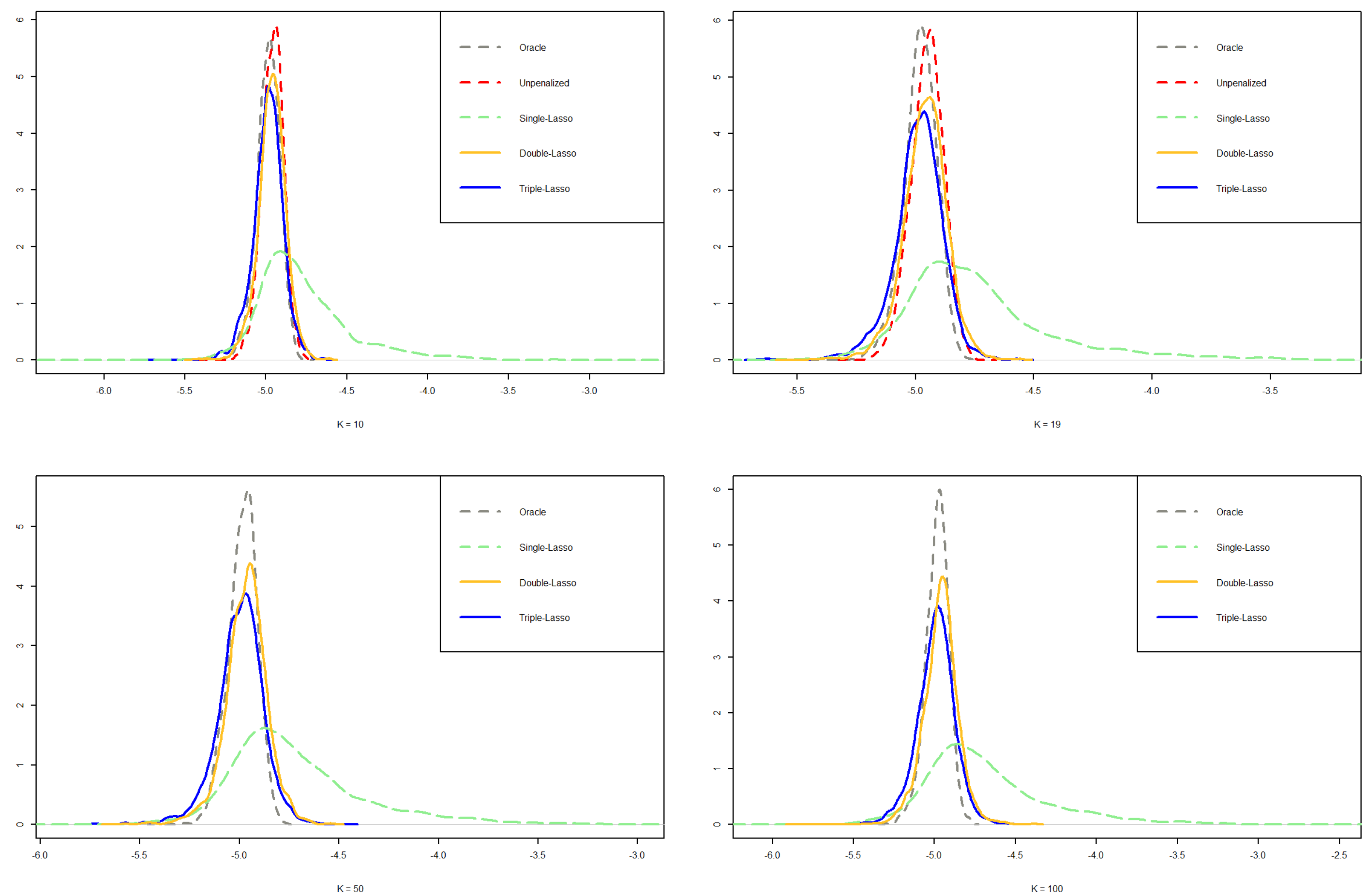
	Z1	Z2	Z3	Z4	Z5	Irrelevant(each)
Relevance	0.7	0.49	0.343	0.2401	0.16807	0
Freq. of Z included	100%	30%	21%	20%	22%	25%

The table reports the relevance and how often we include the specified instrument in the model through our selection process. Z1, Z2, ..., Z10 indicate the respective instruments obtained during DGP. 'Irrelevant(each)' reports the information on each of the irrelevant(Z6, ..., Z10) instruments. 'Freq. of Z included' reports how often the respective instrument is included in the model. I used 10-fold cross-validation and choose the penalty term that generates the lowest error.

As I allow correlation among Z variables during DGP, the relevant instruments may show up less frequently than irrelevant instruments. However, it is a nice property that the most relevant Z is almost always included in the model. Thanks to this property, the Triple-Lasso procedure is not distorted after instrument selection.

The density graphs in Figure 1 are also similar to what we observe in Gillen et al. (2014). The Triple-Lasso procedure seems to capture asymptotic normality. For a smaller number of characteristics, the results are sufficient for all estimators except Single-Lasso. As previous results suggest, for high dimensional characteristics Triple-Lasso and Double-Lasso perform well with no clear superiority over each other.

Figure 1: Density graph for β_p



The figure plots the density of the price coefficient estimated by respective models. These models are described in 3.2. I analyzed the performance in four different cases $K \in \{10, 19, 50, 100\}$. As Unpenalized 2SLS is not defined for more than 19 characteristics, it is not included for $K \in \{50, 100\}$.

Table 4: Sampling Properties of Estimated Price Parameters for Different Number of Instruments

	K = 10			K = 50		
	L = 10	L = 20	L = 50	L = 10	L = 20	L = 50
Bias for β_p						
Oracle Benchmark	0.015	0.016	0.018	0.017	0.017	0.016
Unpenalized 2SLS	0.041	0.079	0.146	-	-	-
Post-Triple Lasso	0.012	0.026	0.048	0.004	0.022	0.043
Post-Single Lasso	0.236	0.346	0.528	0.285	0.440	0.667
Post-Double Lasso	0.037	0.079	0.148	0.034	0.081	0.154
Standard error for β_p						
Oracle Benchmark	0.073	0.071	0.072	0.074	0.071	0.071
Unpenalized 2SLS	0.069	0.061	0.049	-	-	-
Post-Triple Lasso	0.097	0.092	0.093	0.120	0.112	0.114
Post-Single Lasso	0.306	0.322	0.383	0.347	0.372	0.437
Post-Double Lasso	0.088	0.077	0.060	0.109	0.089	0.075
Root Mean Squared Error for β_p						
Oracle Benchmark	0.074	0.073	0.074	0.076	0.073	0.073
Unpenalized 2SLS	0.080	0.099	0.154	-	-	-
Post-Triple Lasso	0.098	0.096	0.104	0.120	0.115	0.122
Post-Single Lasso	0.386	0.472	0.652	0.449	0.576	0.797
Post-Double Lasso	0.095	0.110	0.160	0.114	0.120	0.171

The table reports the performance of different models for a different number of characteristics and instruments. I used $K \in \{10, 50\}$, and for each of these choices, I simulate the results for a different number of instruments, $L \in \{10, 20, 50\}$. The properties are the same as the simulation used in Table 1. I set $c_1 = 0.7$, $c_2 = 0.5$, I keep the number of relevant instruments fixed at 5, and simulate the result 2500 times for each case.

So far, the results do not show a clear benefit of the third step. Actually, for some cases, we observe a slightly better performance of the Post-Double Lasso procedure. This result suggests that omitting the third step may be a better choice. To observe the impact of the third step explicitly, I increase the number of instruments, and the results are summarized in Table 4. For each $K \in \{10, 50\}$ a different number of instruments, $L \in \{10, 20, 50\}$, are generated and the performance of the respective estimator is reported. I keep the number of relevant instruments fixed at 5.

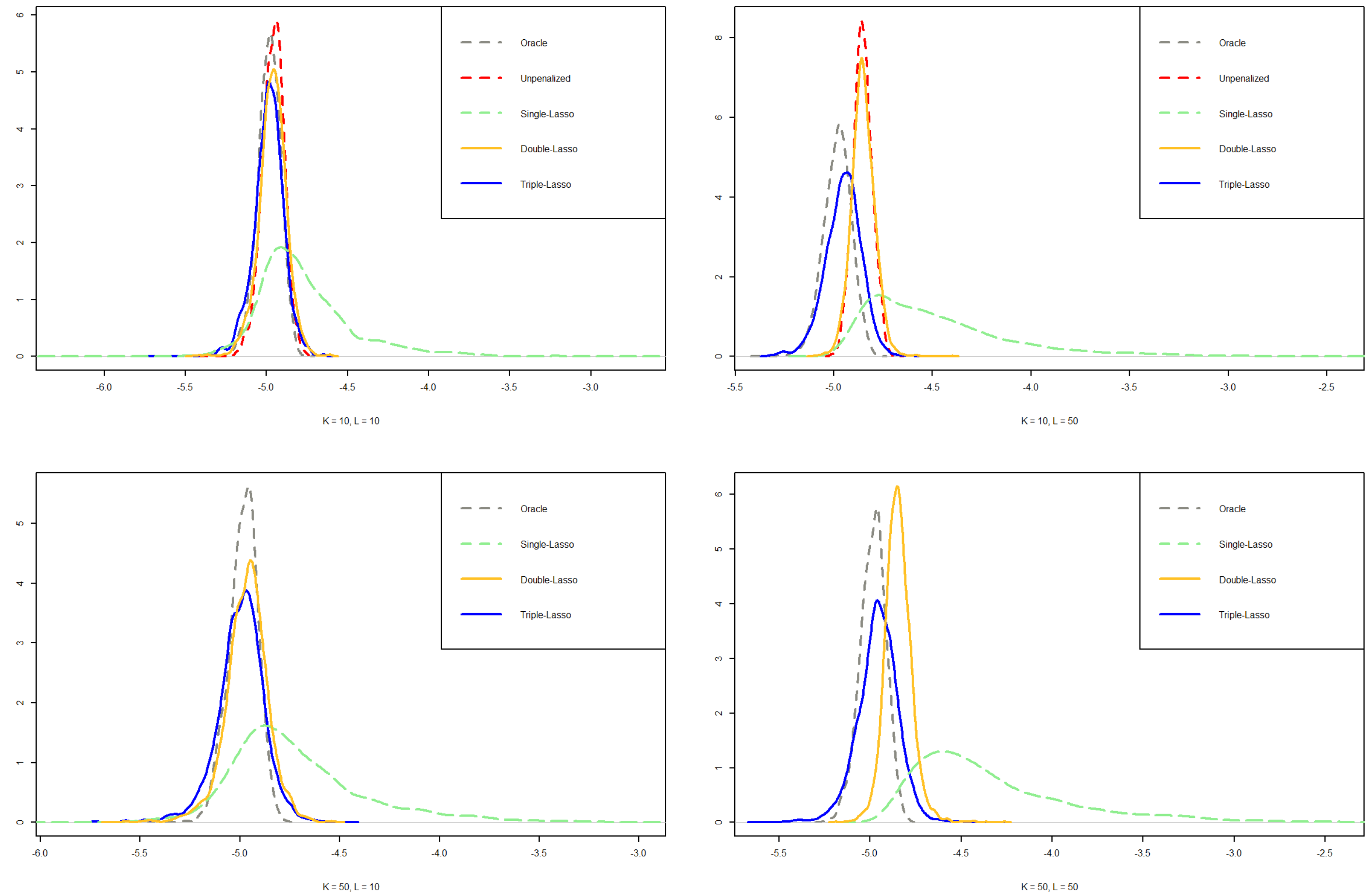
As expected, increasing the number of instruments decrease the standard errors keeping everything else fixed. Surprisingly, for Post-Single Lasso, the opposite is observed. Probably, selecting relevant characteristics erratically result in obtaining more distant coefficients in each simulation. This behavior may need a closer inspection.

I observed a significant increase in the root-mean-squared error for the Post-Double Lasso procedure which omits the instrument selection step. The increase in RMSE is derived from the increase in bias as it exceeds the decrease in standard error. The best results were obtained with Post-Triple Lasso even though its selection of relevant instruments is not perfect. When I use 50 instruments, on average around 5.5 instrument is selected in total, and the most relevant instrument is selected perfectly. The selection pattern is similar to 3 with lower percentages except for Z1.

As we observed in Figure 1 when the number of instruments is small, in figure 2, there is no clear superiority between Double-Lasso and Triple-Lasso. However, when we increase the number of instruments to 50, the Double-Lasso estimator has a significant bias. The mean of the Post-Double Lasso procedure is more than two standard deviations away from the real price coefficient. Such a result may cause wrong interpretations. So, including the third step is crucial for high-dimensional instruments.

Up to now, I generated instruments that have a sufficient correlation with the price. However, in a real setting, we may obtain weak instruments. I will simulate similar results with weak instruments. Previously the correlation of price with the most relevant instrument was 0.7, now I decrease this number to 0.2. The correlation is even lower for subsequent instruments. Table 5 summarizes the same simulations with Table 4 only for the weak instrumental variables.

Figure 2: Density Graph of β_p Controlling for the Number of Instruments



The figure plots the density graph of β_p obtained from different estimators. I have controlled the performance of estimators for a different number of characteristics and instruments. I used $K \in \{10, 50\}$ and for each number of characteristic I controlled for two different number of instruments $L \in \{10, 50\}$. DGP is described in 3.1. I used $c_1 = 0.7$, $c_2 = 0.5$, I keep the relevant instruments fixed at 5. Only irrelevant instruments are added when increasing the number of instruments.

Using weak instruments, the performance of all estimators decreases significantly. Even 'Oracle Benchmark' has much higher Bias and Standard Errors. For Post-Triple Lasso and Post-Single Lasso, increasing the number of instruments from 10 to 50, I observe a more acceptable Mean Squared Error. The decrease mostly resulted from the variance. This kind of change may have a harmful effect leading wrong interpretation for the price coefficient. Another important finding with weak instruments is that the advantage of using Post-Triple Lasso has been lost. We experience similar results in both cases for high-dimensional instruments with an obvious superiority of the Post-Double Lasso procedure when the number of instruments is lower. These results suggest that, if we have weaker instruments, omitting the third step may be a better idea.

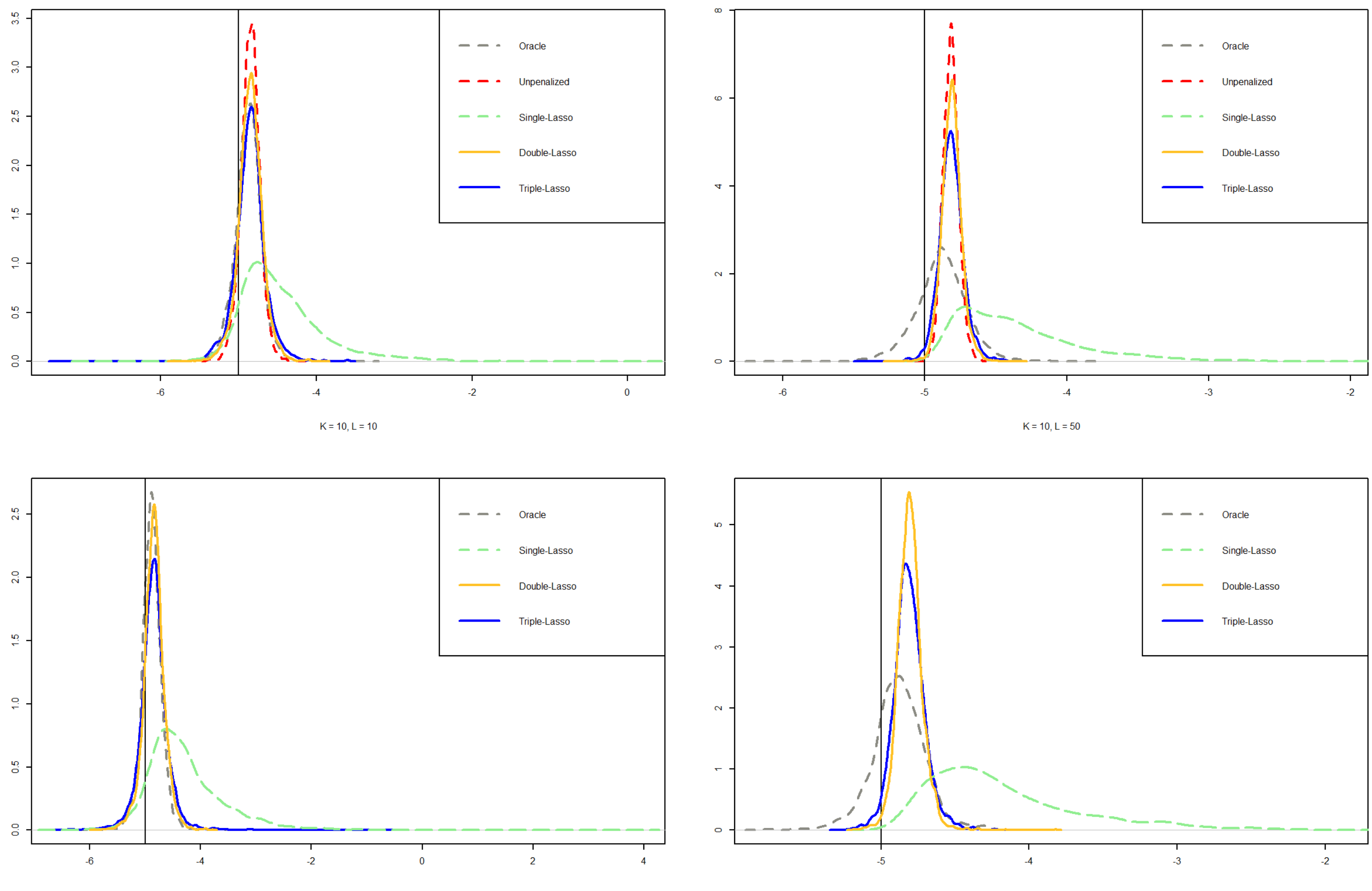
Table 5: Sampling Properties of Estimated Price Parameters for Different Number of Instruments for Weak Instruments

	Bias for β_p					
	K = 10			K = 50		
	L = 10	L = 20	L = 50	L = 10	L = 20	L = 50
Oracle Benchmark	0.120	0.118	0.119	0.114	0.123	0.120
Unpenalized 2SLS	0.151	0.169	0.183	-	-	-
Post-Triple Lasso	0.146	0.170	0.186	0.153	0.180	0.189
Post-Single Lasso	0.551	0.620	0.650	0.722	0.759	0.822
Post-Double Lasso	0.152	0.176	0.189	0.159	0.184	0.195
	Standard error for β_p					
Oracle Benchmark	0.176	0.184	0.188	0.181	0.192	0.183
Unpenalized 2SLS	0.125	0.089	0.056	-	-	-
Post-Triple Lasso	0.212	0.143	0.089	0.285	0.175	0.109
Post-Single Lasso	0.564	0.504	0.441	0.709	0.563	0.515
Post-Double Lasso	0.161	0.119	0.071	0.196	0.141	0.088
	Mean Squared Error for β_p					
Oracle Benchmark	0.213	0.218	0.222	0.214	0.228	0.219
Unpenalized 2SLS	0.196	0.191	0.192	-	-	-
Post-Triple Lasso	0.258	0.223	0.206	0.323	0.251	0.218
Post-Single Lasso	0.789	0.799	0.785	1.012	0.945	0.970
Post-Double Lasso	0.222	0.213	0.202	0.252	0.232	0.214

The table reports the performance of different models for a different number of characteristics and instruments. I used $K \in \{10, 50\}$, and for each of these choices, I simulate the results for a different number of instruments, $L \in \{10, 20, 50\}$. The properties are the same as the simulation used in Table 1. I set $c_1 = 0.2$ to observe the performance with weak instruments and $c_2 = 0.5$. I keep the number of relevant instruments fixed at 5, and simulate the result 2500 times for each case.

Figure 3 makes it clear that for the lower number of instruments, improvement is not likely since the estimators perform similarly to Oracle Benchmark. For a higher number of instruments, however, Oracle Benchmark has a lower bias. For Double-Lasso and Triple-Lasso procedures, the variance decreases with the increasing number of instruments causing wrong interpretations for the price coefficient. The percentage of confidence intervals that cover the real price coefficient decrease from 87% to 48% for the Triple-Lasso and 82% to 30% for the Double-Lasso Procedure. The improvement is only possible with a better selection of instruments.

Figure 3: Density Graph for β_p with Weak Instruments



The figure plots the density graph of β_p obtained from different estimators. I have controlled the performance of estimators for a different number of characteristics and instruments. I used $K \in \{10, 50\}$ and for each number of characteristic I controlled for two different number of instruments $L \in \{10, 50\}$. DGP is described in 3.1. I used $c_1 = 0.2$ to observe the performance with weak instruments and $c_2 = 0.5$. I keep the relevant instruments fixed at 5. Only irrelevant instruments are added when increasing the number of instruments.

Table 6: Sampling Properties of Estimated Price Parameters for Different Number of Instruments and Penalty Terms

	Bias for β_p					
	L = 10			L = 50		
	λ = 0.1	λ =0.2	λ =0.3	λ =0.1	λ =0.2	λ =0.3
Oracle Benchmark	0.041	0.041	0.041	0.039	0.039	0.039
Unpenalized 2SLS	-	-	-	-	-	-
Post-Triple Lasso	0.047	0.006	0.023	0.245	0.142	0.033
Post-Single Lasso	0.368	0.368	0.368	0.795	0.795	0.795
Post-Double Lasso	0.091	0.091	0.091	0.313	0.313	0.313
	Standard error for β_p					
Oracle Benchmark	0.114	0.114	0.114	0.113	0.113	0.113
Unpenalized 2SLS	-	-	-	-	-	-
Post-Triple Lasso	0.158	0.175	0.181	0.106	0.135	0.168
Post-Single Lasso	0.376	0.376	0.376	0.417	0.417	0.417
Post-Double Lasso	0.141	0.141	0.141	0.092	0.092	0.092
	Root Mean Squared Error for β_p					
Oracle Benchmark	0.348	0.121	0.121	0.120	0.120	0.120
Unpenalized 2SLS	0.152	0.152	0.152	0.324	0.324	0.324
Post-Triple Lasso	0.165	0.175	0.182	0.267	0.196	0.171
Post-Single Lasso	0.526	0.526	0.526	0.897	0.897	0.897
Post-Double Lasso	0.168	0.168	0.168	0.327	0.327	0.327

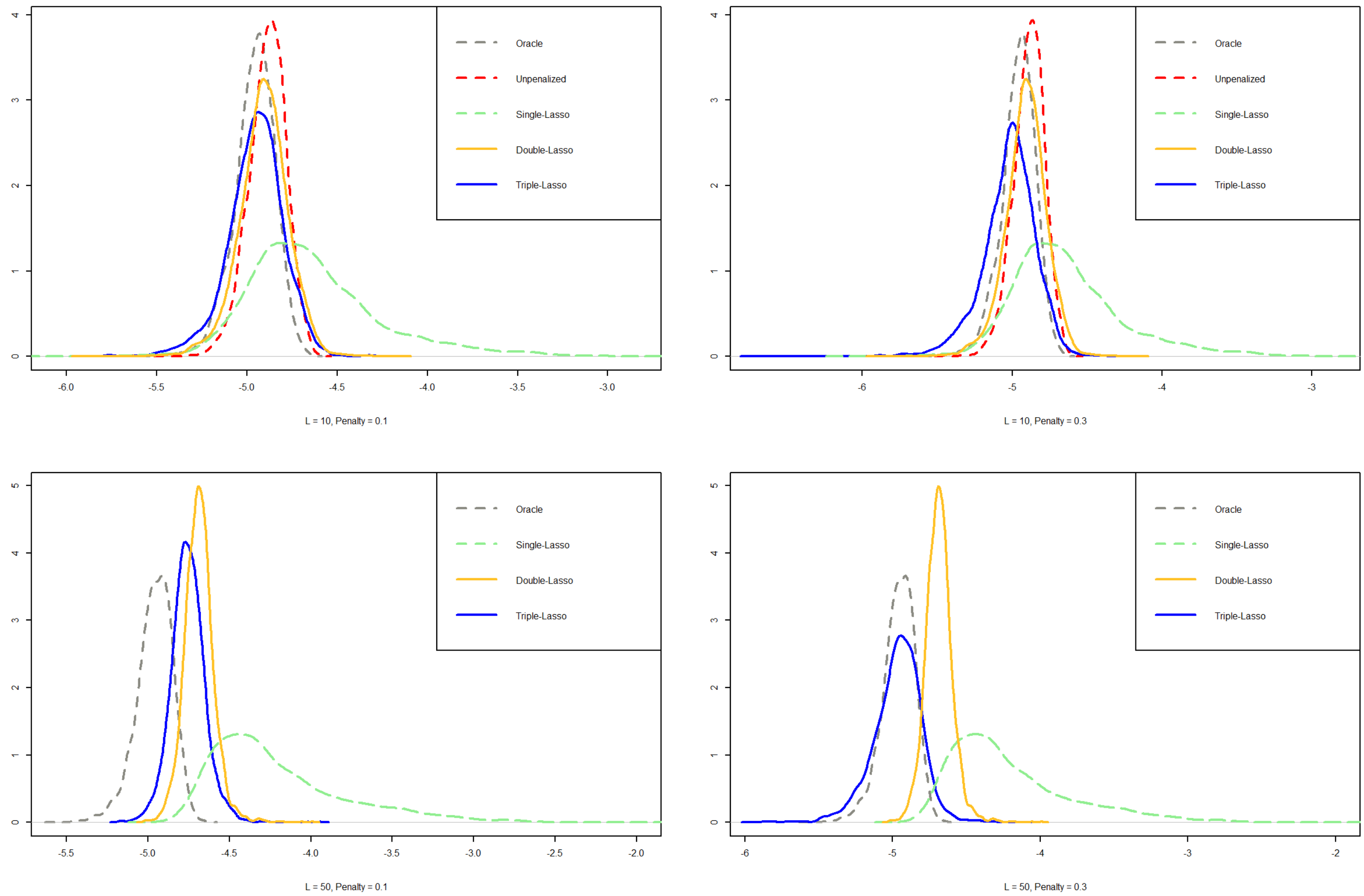
The table reports the performance of different models for a different number of instruments and penalty terms. I keep K fixed at 10. I control for instruments, $L \in \{10, 50\}$. For both case I use different penalty terms, $\lambda \in \{0.1, 0.2, 0.3\}$. The DGP is described in 3.1. I set $c_1 = 0.7$ and $c_2 = 0.8$ to increase the intensity of endogeneity. I keep the number of relevant instruments fixed at 5 and simulate the result 2500 times for each case.

In the previous analysis, I used a moderate level of endogeneity by setting c_2 to 0.5. Changing c_2, we can manipulate the intensity of endogeneity. Intuition says that if we increase the intensity of endogeneity, we expect to obtain even worse performance in the case of weak instruments. To observe this, I increased the endogeneity by changing c_2 to 0.8. The performance for each estimator was worse than Table 5 as expected. As an improvement strategy for selection, I control for the penalty term but no significant improvement has been observed.

With sufficiently correlated instruments, we obtain good predictions using the Triple-Lasso procedure even with stronger endogeneity. The importance of instrument selection became even more explicit with more intense endogeneity. The bias is more than three standard errors for the most extreme Post-Double Lasso case which may cause extremely flawed interpretations. Only in 8.5% of the cases, the confidence interval for Post-Double Lasso covers the true price coefficient while it is 74% for the Triple-Lasso procedure. In Table 6, we control performance for stronger endogeneity with different penalty terms. Again, with a lower number of instruments, no superiority between Double-Lasso and Triple Lasso is observed. We seem to have also more flexibility when choosing the penalty term. For $L = 50$; however, higher penalties are favorable. We know that for consistent selection, $\lambda/N \to 0$ and $\lambda/\sqrt{N} \to \infty$ should hold. The results in Table 6 suggest that in finite cases it is a good idea to further adjust the penalty term depending on the dimension.

The Figure 4 make it clearer that for high-dimensional instruments a smaller penalty term is not sufficient even though it performs well for lower number of instruments. Overall, Triple-Lasso procedure performs well with the right penalty term when we have sufficient correlation between instruments and price regardless the size of endogeneity.

Figure 4: Density Graph for β_p with Different Penalty Terms



The figure plots the density graph of β_p obtained from different estimators. I have controlled the performance of estimators for a different number of instruments and penalty terms. I keep K fixed at 50. I used $L \in \{10, 50\}$ and for each number of instruments I controlled for the penalty term $\lambda \in \{0.1, 0.3\}$. DGP is described in 3.1. I used $c_1 = 0.0.7$ and $c_2 = 0.8$ to increase the intensity of endogeneity. I keep the relevant instruments fixed at 5. Only irrelevant instruments are added when increasing the number of instruments.

4 Conclusion

I replicated the procedure offered by Gillen et al. (2014) using fixed instead of random coefficients. The results I obtained are similar to Gillen et al. (2014). I got better performance in terms of Sampling Properties for β_p due to decreasing variance caused by fixed coefficients. However, the performance for characteristics selection was slightly worse, probably due to different selection techniques. In addition to the analysis in Gillen et al. (2014), I put a strong emphasis on the instrument selection step. I showed that including every instrument in the model is not a great idea and causes flawed interpretations. The selection of instruments becomes even more crucial if we have more intense endogeneity. For a weaker and lower number of instruments, however, the superiority of the Triple-Lasso procedure became ambiguous.

Further analysis would involve trying to employ a better selection technique for the relevant instruments and characteristics. We may achieve a better selection with a proper adjustment in the penalty term benefiting from the literature. Potential techniques that could bring easy improvements in the selection could be Adaptive Lasso, Elastic Net, etc.

Starting from Chamberlain (1987), a large body of literature is discussing optimal instruments. The effect of optimal instruments on BLP models is discussed by Reynaert and Verboven (2014). Adding additional steps to characterize the optimal instruments suggested by the literature would also be an interesting topic.

Additionally, in my future work, I want to focus on Random Coefficient Model and its variants. Specifically, Analysing the importance of instrument selection, and the effect of employing optimal instruments on models such as Dynamic Demand Estimation Hendel and Nevo (2006), and Nonparametric Identification Techniques Berry and Haile (2014) would be interesting.

References

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies*, 81(2):608–650, 11 2013. ISSN 0034-6527. doi: 10.1093/restud/rdt044. URL https://doi.org/10.1093/restud/rdt044.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica*, 63(4):841–90, 1995. URL https://EconPapers.repec.org/RePEc:ecm:emetrp:v:63:y:1995:i:4:p:841-90.

Steven T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262, 1994. ISSN 07416261. URL http://www.jstor.org/stable/2555829.

Steven T. Berry and Philip A. Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014. doi: https://doi.org/10.3982/ECTA9027. URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA9027.

Gary Chamberlain. Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334, 1987. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(87)90015-7. URL https://www.sciencedirect.com/science/article/pii/0304407687900157.

Benjamin J. Gillen, Matthew Shum, and Hyungsik Moon. Demand estimation with high-dimensional product characteristics. In *Bayesian Model Comparison*, volume 34, pages 301–323. Emerald Publishing Ltd, 2014. URL https://EconPapers.repec.org/RePEc:eme:aecozz:s0731-905320140000034020.

Igal Hendel and Aviv Nevo. Measuring the implications of sales and consumer inventory behavior. *Econometrica*, 74(6):1637–1673, 2006. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/4123086.

Mathias Reynaert and Frank Verboven. Improving the performance of random coefficients demand models: The role of optimal instruments. *Journal of Econometrics*, 179(1):83–98, 2014. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2013.12.001. URL https://www.sciencedirect.com/science/article/pii/S0304407613002649.