Poverty in Latin America: theory and statistical application with lasso regression

Alejandro-Matias Gonzalez Ferraro

Code: EDE05
Tutor: Piotr Zwiernik
2015/2016

Abstract

Poverty has been studied from many different points of view based on both historical data and theoretical approach. We take a statistical approach, and economic theory is used to both select the group of variables taken into account and to interpret the results and the validity of the econometric model.

In this paper we discuss the *lasso* regression, which is a statistical tool to obtain sparse solutions to regressions problems. Lasso regression has many applications, from biology to economics. Our proposed application is in social economics, specifically in the analysis of poverty rate determinants. The goal is to find variables which have higher effect on poverty. For this specific case study we focused on Latin American countries which have lower level of income and high poverty headcount ratio. The statistical results obtained where as expected by the theoretical framework, and the different statistical specifications conclusions are similar.

Acknowledgements

I want to give my gratitude to professor Piotr Zwiernik for the help and attention on elaborating this final year project.

Contents

1 Introduction

In this research report we focus on the analysis of poverty in a group of Latin American countries. We want to explore statistically the lower performance countries and show both whether economic theory and intuition are consistent with the data and try to infer some results which may be useful from public economics point of view and to future applications of *lasso*.

Lots of studies have been made about this topic [1–3], and we will quickly review a few of them. This will help us to pre-select the variables for our model. We discuss how important these variables are in terms of explaining poverty rates in Latin American countries.

We use the *lasso* regression to identify in the pre-selected set a small subset of the most important poverty determinants. *Lasso* regression also fist well our context where we have few observations and a lot of covariates. A generalization of *Lasso* elastic net is also used to put more stable estimates in the case when some variables might have correlation between them, which happens also in our case.

In our statistical model we considered time series data from Colombia, Paraguay, Brazil, Costa Rica, Ecuador, Panama and Peru (the country and the last 3 years). These where the ones that had larger quantity of data among the poorest countries. Furthermore, we wanted to pick countries from different regions of South and Central America. We expect the parameters to be similar because they are in the group of the poorest Latin American countries.

This paper is organized as follows. In Section 2, we review the methodology of poverty studies in general and in Latin America. In Section 3, we introduce *lasso* regression and the concept of *sparsity*, with the purpose of showing why it can be applied in our context. Section 4, shows the methodology for the computation of the model. In Section 5 we interpret the results, compare the models, and make some conclusions from the economical and social point of view. In Section 6 we present the conclusions of our research.

2 Poverty in Latin America

2.1 Socio-economical context

Latin America, in the 1970's had a difficult period because of the oil crisis. In this period, both imports and the foreign market relations decreased. Increased debt service to foreign investors caused social aids to decrease. There were also currency devaluations which worsened national consumption and foreign investment [4].
In recent decades, Latin American countries performance in terms of growth has improved [5]. However, there is still a major problem, which requires taking special attention by governments and international institutions, that is, poverty rate and income distribution.

Latin American countries have long been suffering from a bad macroeconomic policy which led them to financial disasters in terms of sovereign bond crises. Moreover, corruption has deteriorated international relations and money management. In 1981, more than one third of Latin America's population had income below the poverty line, that is, with income below $1.90 a day in purchasing power parity terms[8]. The debt crisis from 1982 worsened the situation and also there was still a significant quantity of children malnutrition [2]. Even if the tendency seemed to reverse in recent years for some countries, there are still others which need to take public policy action.

Some of the countries have an income per capita that is sufficiently high for helping improve social conditions. However, there are still high poverty rates in the society. For example, people who live below the poverty line in Brazil is around 35% of the total population. *Income distribution*, is one of the problems which can explain this, 1% of the population holds 14% of the total income in Brazil [7]. Other countries like

Peru, have serious problems with education [7]. Higher education depends on social status, so, it is far more difficult for the poorest to get access to it, decreasing the opportunities to get out of poverty.

2.2 Previous research on poverty

Our research differs from past studies in the sense that we used theoretical framework to apply a simple interpretable statistical model that can be used as a filter to focus on certain groups of factors that affect poverty. In this subsection, we briefly present some of the related literature and results of previous poverty studies in Latin America. When studying importance of education in Latin American countries [3, Section 3], *logarithmic regression*, is used when we are interested in returns of the objective variable or growth rates, the only difference with standard least squares regression is that the objective variable is a logarithmic transformation of the original one. This is often applied to show the effect of education in earnings, which implies a reduction on poverty:

$$lnY_i = \alpha + \beta S_i + \phi_1 E_i + \phi_2 E_i^2 + \epsilon_i, \tag{1}$$

where α is the intercept, S is the number of years of schooling of individual i, and E represents years of experience. The square of E is used to represent the diminishing marginal returns.

Focusing on earnings effects to predict poverty can be misleading, people can go to school but at the same time they are working and, due to that, school years do not have the same effect on children from poorer families as for others. Thus, education effect on fighting poverty can be *underestimated*.

Other studies [1], tend to use data for *descriptive analysis* which refers to the study of the first two moments of the distribution. There is also an interesting publication that highlight the role of capital flows [7], that is, the reduction on inflation rates and the fiscal balances improved, but there are still many problems concerning macroeconomic stability and searching of foreign markets to invest.

Previous studies rely on the variables which had more data available, the poverty data were not as developed as it is now. We try to overcome that problem using the most recent database on poverty from World Bank Database. A lot of poverty indicators are available and used in previous studies. We have chosen poverty headcount ratio at $3.10.

2.3 Factors of Poverty in Latin America

Macroeconomic stability was an important factor in explaining poverty [1, 4, 7]. The deficit in the balance of payments causes inflation and tax growth which directly affects the poorest, specially if there is a non-efficient tax system and corruption, which is the case of some Latin American countries [7]. Furthermore, deterioration in the balance of payments also affects the export dependent enterprises and so the labor market and wages of the workers.

Increasing GDP (Gross Domestic Product) growth in a country is one of the most widely known factors that contribute to reducing poverty, but one important point to consider is the *type* of growth in the country. Different types of growth determine how poverty rate is reduced. Important factors to take into account to analyze the type of growth a country is whether there is a direct investment in human capital or not, if the firms market is national or international. However, the most important is how the benefits of growth are distributed among the population [1].

Poverty studies are also related to *health and population* [2]. One of the problems which we observed is related to the focus on *curative* health rather than on *preventive* health. Curative health, can be a contra-productive measure as it emphasizes more on purchase of drugs and medicines which is a more expensive and inefficient way of targeting the health problem. The factors which determine the uses of these two can be related both to government expenditure on health and education of the population.

A recently published book [6, pp 54-56] mentions one important factor to take into account, the quality of the *government health expenditure*. In other words, it is not the quantity of money invested by government on health that matters but *how* they are invested.

Previous studies [3, 7] agree that *education* is a relevant factor in explaining poverty in Latin America, at least considering the decades of the 60's and 70's. *Education* in Latin America has problems with the high quantity of students repeating courses, together with the enrollment rates and the quality of the education implemented [3]. This repetition implies high cost for subsiding education. The quality of education is fundamental to have the labor force adapted to globalization and competition between countries. We can learn from International Economics Theory [9] that the human capital characteristics of the country determine whether it specializes in one product or another. Generally, countries with more human capital will tend to produce products with high value added and, due to that, the richer the economy and population will be. The fundamental problem with this specialization is that there is no solution for the latter apart from a structural change taken by the government or the households investment in education so these are two things to take into account when studying the future performance of human capital.

Fertility rate, can in many ways be related with poverty rate. For example, larger poor families, will have less resources for the education of their children. Then, instead of studying, the children may study and work or just simply work, affecting negatively the performance and the returns from education. This fact widens the gap between the richest and the poorest [6, pp 81-86].

Nutrition and *social security* plans taken by the government are also an important issue to consider, one reason is because they directly affect the performance of the children at school and due to that the chances of getting away from poverty. This *social security* programs are also inefficient in terms of effectiveness and optimization of costs [7].

3 Sparsity and lasso

We live in times when the quantity of databases is raising at an unexpected level compared to past decades. So there is a need for the use of computational models which can deal with such an abundance of data. The idea behind a *sparse* regression is the assumption that many variables in the regression are associated to zero parameters. When $p > n$, where p is the number of explanatory variables and n is the number of observations, this happen in cases when you have a large quantity of variables that you think are important in defining your objective variable, and you have less observations. In this case, *ordinary least squares estimator* is not well-defined because of the mathematical properties of the matrix it assumes. You can try to exclude the variables yourself by simply trying to discriminate using previous studies or theoretical framework. However, this method is lacking objectivity and effectiveness.

In cases where $n \geqslant p$, to reduce the number of covariates in the regression, you can do a two step procedure using *least squares estimation*, which consists in regressing the vector of variables x_ij against the objective variable y_i and then excluding those with less explanatory value. Then, try to fit the new model again with the remaining explanatory variables. But this will imply difficulties in both computation and *accuracy* of the model. There are other types of computational tools that are also adequate in the context where we have a lot of explanatory variables, but in our paper we will consider only *elastic net*, which together with *lasso* we discuss in more detail in the next section.

3.1 Lasso regression

Suppose we are interested in which variables x affect y and we assume that the relation between them is linear. In this case we can simply apply *linear regression*:

$$y_i = \beta_0 + \sum_{j=1}^{p} x_{ij}\beta_j + \epsilon_i, \tag{2}$$

where β_0 and $\beta = (\beta_1, \beta_2, ..., \beta_p)$ are unknown parameters, x_{ij} is the i^{th} observation of the j^{th} covariate, and ϵ_i is the error. We assume that the errors are independent and identically distributed.

The most popular way of estimating the parameters of (2) is by using the *least squares estimator*:

$$\min_{\beta_0,\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{3}$$

However, when the number of explanatory variables to fit grows, it is difficult to get a proper fitted model without increasing the variance of the estimated parameters, and if $p > n$ the model is not well defined. This happens, for example, in cases of machine learning or in health economics where you have a lot of potential variables determining your objective variable. When dealing with this type of data the *lasso* works better.

Mathematically the *lasso* regression [12] is obtained by:

$$\min_{\beta_0,\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta j)^2 \qquad \text{subject to} \quad \|\beta\|_1 \leqslant t \tag{4}$$

where $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$ is the ℓ_1 norm of β and $t > 0$ is a fixed number. We use the ℓ_1 norm because makes calculations *easier* in mathematical terms and it is a simple optimization problem and at the same time leads to sparse solutions [10, Section 2.2].

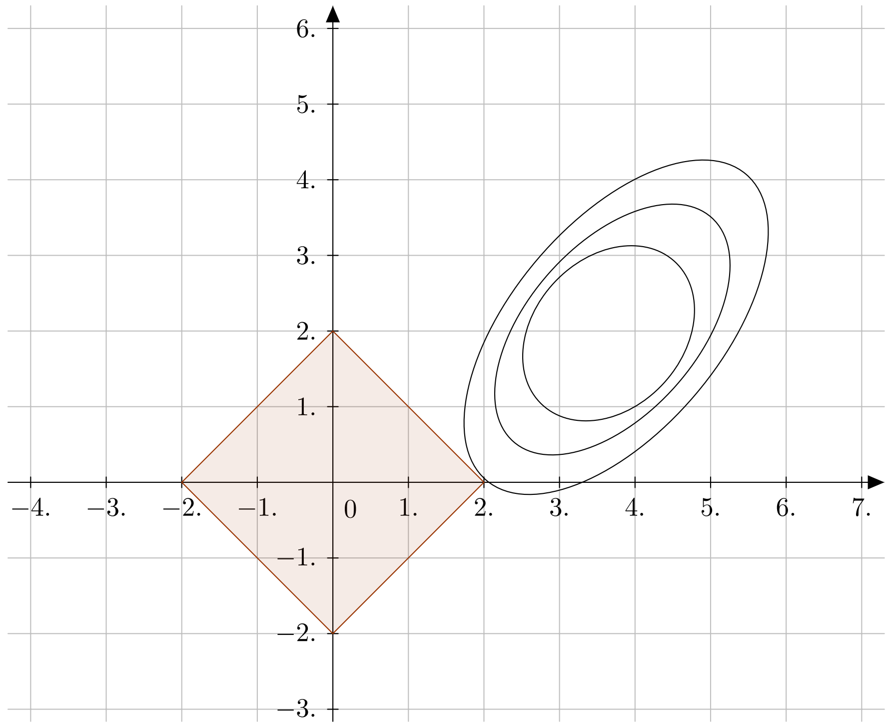


Figure 1: Lasso optimization

The shaded region is the constraint $|\beta\|\leqslant 2$. The ellipses are the level sets of the function to optimize. The optimal solution for the *lasso* regression will be found on a corner because we are minimizing a strictly convex function subject to an absolute value (the rotated square), $|\beta|$. The consequence is that we will have some parameters β that will have zero value of.

We have presented one mathematical form of representing *lasso* regression (4). Alternatively, we can see it as a kind of Lagrangian problem:

$$\min_{\beta\in\mathbb{R}}\left\{\frac{1}{2N}\|y-x\beta\|_2^2+\lambda\|\beta\|_2\right\} \tag{5}$$

where y is the vector containing the responses of the objective variable for each observation n, and x are the explanatory variables. Formulations (4) and (5) are equivalent, for a particular value of λ.

The interesting thing about this mathematical approach is that the values of λ, the penalization parameter, determine which coefficients will have zero value. In this context, we should use a different computational approach because the function is not differentiable, that is, we cannot use the partial derivatives to find the optimum. In this context, we use the *subdifferential*.

The *coordinate descent method* is an algorithm that solves *lasso* problem by fixing all coordinates but one, leading to a one variable optimization problem in every step until convergence is reached. If we minimize the function for every coordinate it can be proved that if the function is minimized along each coordinate we can reach a global optimum [10, Section 2.4.2].

3.2 Cross-validation and forecasting

The value of t which bounds the parameter β in (4) can change the fit of the data in our model. On one hand, if you increase the bound t the more parameters you will have in your model as *restriction* decreases and the goodness-of-fit of the model increases. On the other hand, if you decrease the bound t the less parameters you will have and also less fit in the data, but you may gain the model *interpretation*. For example, it is not the same to have to find the relation of three explanatory variables and our objective variable, than have to interpret 38.

There is a direct way of finding the best value for t in terms of mean squared error, and it is using *cross-validation*. The procedure is as follows. Randomly divide your dataset say, in two groups, and then try to predict one with the other for different values of t, you could also divide it in more than two, in fact, this is what cross validation does in our R computations. Then, the *cross validation error curve*, which informs what is the minimum mean squared error for different values of t is generated. Finally, we will select the t which gives us the minimum mean squared error, which is: $\text{MSE} = \frac{1}{n}\sum_{t=1}^{n}\omega_t^2$, where ω_t^2 is the squared difference between your prediction and the real value.

3.3 Elastic net

In this section we introduce the *elastic net*. There we combine the ℓ_1 penalty and the ℓ_2 penalty, which generally performs well in contexts where we have correlation in our explanatory variables. The *lasso* does not work well when we have correlated variables, in the extreme case where we have two variables which are equal, *lasso* gives the same weight to both. That is when we should consider to use *elastic net* which groups the variables [10, Section 4.2].

As we stated before, the *elastic net* is obtained by using a combination of the two penalties, mathematically:

$$\min_{(\beta_0,\beta)\in\mathbb{R}\times\mathbb{R}^p}\left\{\frac{1}{2}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda\left[\tfrac{1}{2}(1-\alpha)\|\beta\|_2^2+\alpha\|\beta\|_1\right]\right\} \tag{6}$$

where, $\alpha\in[0,1]$. When $\alpha=1$ we obtain the *lasso* penalty.

4 Methodology and data selection

In this section we will explain first the tools we used to obtain the statistical results, second the variable selection process, third computation of the statistical model, and finally how we can infer the results of our model.

4.1 Databases and software used

For this paper we used the World Bank database [8], both for reliability and the large quantity of data. We performed our computations in R. We have chosen R both for the computation easiness and the quantity of packages that are available for our specific statistical studies. For *lasso* computation in R we used the package called *glmnet* [13], which is easily downloaded from R-CRAN library. The command *glmnet* fits *lasso* regression using *coordinate descent method*.

4.2 Variables Selection

Care is needed in the variable selection process, because it can be a problem of correlation between variables, that usually leads to *spurious regression*. This happens when two variables are strongly correlated by some exogenous factor not taken into account and due to that fact we can state that one causes the other when it is not the case. The typical example used is that an increase in ice creams sold causes more shark attacks, the external correlation factor here would be the fact that it is summer. By using lasso we can include as many variables we consider in the model but if this variables are correlated the model will select all of them giving similar weights.

We selected the variable poverty headcount ratio at $3.10 a day, because among poverty indicators available, this one had large quantity of data over the years, and also $3.10 a day is poverty but not extreme poverty which will be the case of $1 a day which is another commonly used indicator. GINI coefficient is another commonly used indicator, but it measures structural poverty at a more aggregate level of the country rather than poverty per individual.

The explanatory variables we used in our model can be put into different groups, as discussed in section 2.

4.3 Model Computation

The code we used in our computation of *lasso* regression was:

```
glmmod<-glmnet(x,y=PHR,alpha=1,family='gaussian')
```

This code takes a matrix of explanatory variables and the individuals observations x_i as an input and also the objective variable, which in our case is PHR, poverty headcount ratio at $3.10 a day at international prices. Then, we can also specify the value of α, which will determine which computation we are using to estimate the model, in this case is 1 because we used *lasso*. In case of elastic net, it will be a value between 0 and 1. Finally, you can also specify the type of distribution for the error term, which in our case is *Gaussian*. *Glmnet* standarize the predictors so that each column of the matrix has mean zero and variance 1, because the predictors can be in different units. Cite sec 2.2 sparsity book.

We also computed *cross-validation* of the model, using the following code:

```
1- cv.glmmod <- cv.glmnet(x,y=PHR,alpha=1)

2- plot(cv.glmmod)

3- best_lambda <- cv.glmmod$lambda.min
```

```
4- coef(cv.glmmod, s = "lambda.min")

5- mse.min <- cv.glmmod$cvm[cv.glmmod$lambda == cv.glmmod$lambda.min]

mse.min
```

Line 1 of code computes the values, using a similar input as *glmnet*. Line 2 shows a plot of the mean squared value for every $log(\lambda)$. Line 3 extracts from *cross-validation* computation the λ with the least *mean squared error*. Line 4 shows a table with all the coefficients of each variable. Line 5 extracts from *cross-validation* the *mean squared error*.

5 Poverty application: lasso and elastic net

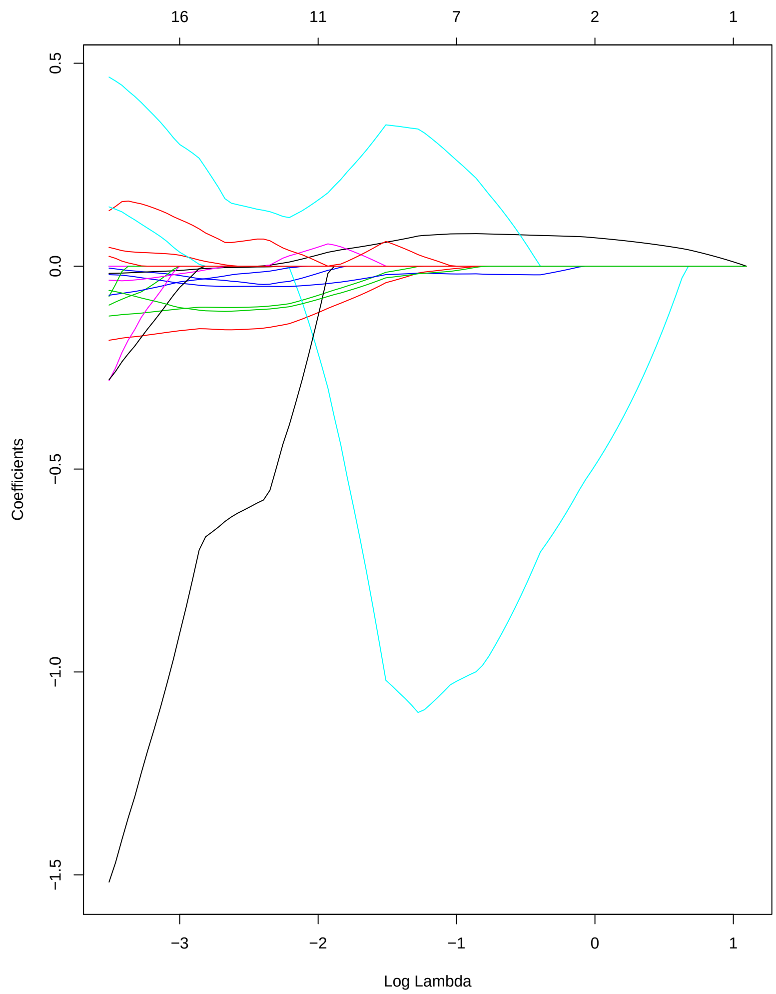
In this section we will estimate the poverty headcount ratio at 3.10$ a day using *lasso* and *elastic net* for computing our statistical model. We used the abbreviation (PHR) for poverty headcount ratio at 3.10$ a day. We considered also using the *logarithmic* transformation of the (PHR) variable, and compare the results in both cases. The explanation of the other parameters abbreviations can be found in the appendix. (Intetar explicar como se interprtan coeff)

We first explain *lasso* computations and results for PHR and log(PHR), followed by the elastic net.

5.1 Lasso

In this subsection we will compute *lasso* first using PHR, and then we will use the *logarithmic* transformation. For computing *lasso* we used the package *glmnet*. The command *glmnet*, shows us a plot which contains important information of our model:

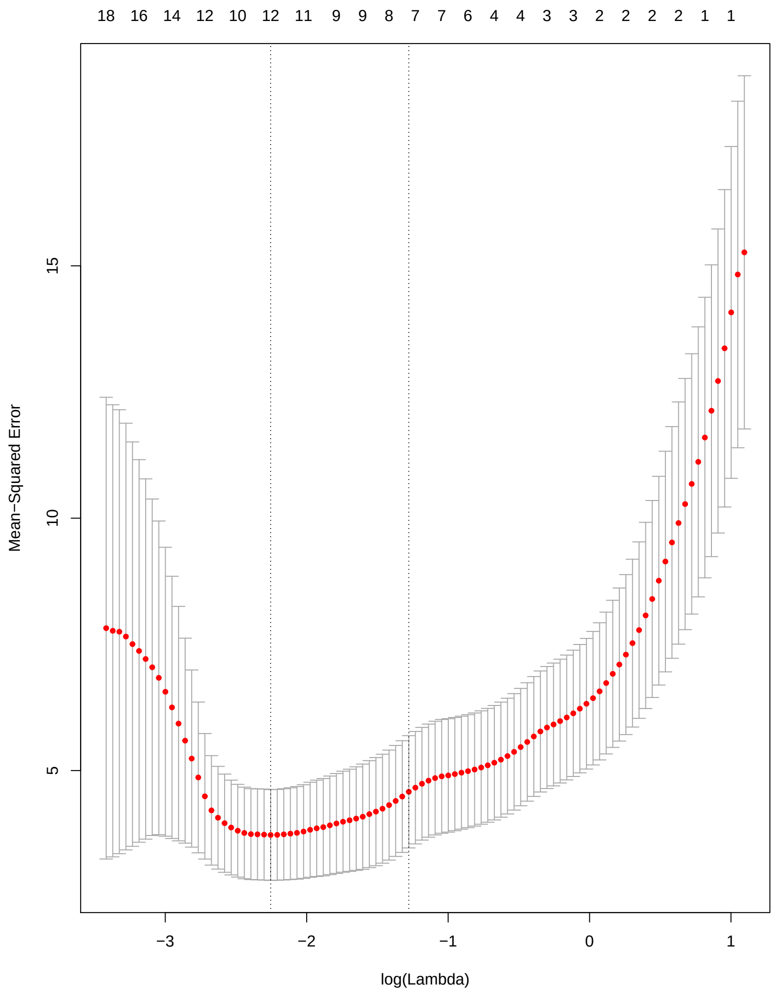
Figure 2: *lasso* coefficients plot



This plot shows the value of the coefficients on the vertical axis and the value of $log(\lambda)$ on the horizontal axis, we also have the number of non-zero coefficients in the model on top. The different curves indicate how the value of the coefficient of each variable changes when we decrease the *penalization*. As we can see, the more we decrease the *penalization* more non-zero coefficients will be in our model.

The plot shows that there are two important explanatory variables for almost all the values of λ. However, depending on the value of lambda we select, they will have different explanatory power. The value of λ we choose for our model will depend on the *mean squared error* of the model, which can be seen by the following plot:

Figure 3: *lasso MSE*



We obtain this plot after computing the *cross-validation* of the model. This plot shows the *mean squared error* on the vertical axis for every value of $log(\lambda)$ on the horizontal axis, on the top we have the number of non-zero coefficients for every value of $log(\lambda)$. We can see that as the value of $log(\lambda)$ increases it also does the *mean squared error*, that is, if we allow for more variables in our model, the error also increases. We must balance the explanatory power of our model with the in sample fit. We can also see that the least *mean squared error* gives us 12 non-zero coefficients.

In the next three following tables, we cannot interpret the value of the coefficients, as they are normalized with mean zero and variance one.

Next, we present a table with the non-zero coefficients values for $\lambda = 0.1832$ corresponding to the least MSE:

Non-zero coefficients & values ($\lambda = 0.1832$)	
Coefficient	Value
FE	0.0075
FOE	-0.1449
ME	-0.0947
SN	-0.0391
AFR	0.01963
CSS	-0.0504
IN	0.1225
EC	-0.0001
ET	-0.4392
EA	0.0450
ES	-0.0066
ESS	-0.1018

The blue coefficients are education parameters and the red ones are exports parameters. The black ones do not have any group in this table.

The table shows different groups. Among those, some important factors are: ET (Employers total% of total employment), FOE (Food exports, % of merchandise exports), IN (Deposit interest rate %) and ESS (Gross enrollment ratio for secondary education).

ET reflects the situation of the *labor market*, the more employers we have, the more employment will be generated and less poverty will be in the economy (negative sign of the coefficient).

FOE represents the situation of international market of the country. Food exports appears as important because some of the countries in our dataset are famous for certain food products (Brazil's coffee for example) [7]. Then, we can conclude that an improve on foreign markets is an important factor for reducing poverty.

IN is an indicator of the financial market. The positive sign of the coefficient that the more deposit interest rate the higher is the poverty rate in the country. This may seem a strange conclusion. A possible explanation is the necessity of liquidity in the economies. An increase on interest rates discourages investments and consumption and due to that, the richness and, consequently, the poverty.

ESS is an indicator for gross secondary education enrollment. It has also a negative sign so an increase in the number of people deciding to study at secondary education is related with reduced poverty.

From this *lasso* analysis, we observe four major areas that influence poverty, labor market, foreign markets, financial markets and education. To target poverty in the poor countries in Latin America it will be recomendable to encourage entrepreneurship, improve the relations with other countries, specifically on food exports, the liquidity of the countries by lowering interest rates and finally investment on education.

Next, we will show the same analysis but taking the *logarithmic* transformation of the PHR variable. We present only the coefficient table, the two plots can be found in the appendix.

Non-zero coefficients & values ($\lambda = 0.0069$)	
Coefficient	Value
FOE	-0.0015
ME	-0.0009
SN	-0.0003
AFR	0.0002
AL	- 0.0002
CEF	0.0005
CSS	-0.0007
ET	-0.1079
EA	0.0010
DSE	4.3558e-16
IN	-0.0042
EC	-0.0001
EE	-0.0016
GEE	0.0017
ESS	0.0010

Were the red ones coefficients represent exports, the green ones labor market, the pink ones financial sector and the blue ones education. The black ones do not belong to any group. We can see that the number of non-zero coefficients increased to 15.

We interpret the coefficients not by its value but by the fact that they are non-zero, that is, all of them predict our objective variable but we cannot say nothing of how much and the sign.

There are some similarities in the coefficients from both regressions: the foreign market, education and labor market relevance, specially the education in the *logarithmic* transformation of PHR.

The value of the MSE for the best λ with PHR is equal to 3.7943, the value for log(PHR) is 0.02794, so, we can see the that the transformation of PHR improves the model considering MSE.

5.2 Elastic net

In this subsection we compute the *elastic net* using *glmnet* command, for $\alpha = 0.5$. We just present the table with the coefficients, plots can be found In the appendix.

Non-zero coefficients & values ($\lambda = 0.3499$)	
Coefficient	Value
FE	0.0046
FOE	-0.0063
ME	-0.0033
SN	-0.0025
AFR	0.0016
CSS	-0.0030
CEF	0.0002
DSE	1.859e-12
IN	0.0248
EE	-0.0719
EA	0.0058
EA	0.0010
ESS	-0.0042

Where the red coefficients represent exports, the purple ones social programs, the green ones labor market and the blue ones education. There is no group for AFR.

Elastic net presents a similar structure as log(PHR), almost the same number of coefficients. Furthermore, the model has the same groups as the two before, there are all taken into account as a group: for example, FE, FOE, ME for exports. The interpretability of this model is more clear, as the model explanation can be done by groups not by variables individually, then we can make more general conclusions. None of the coefficients is more relevant than others, taking into account the λ with the less MSE.

The *mean squared error* with this specification is 4.7811.

5.3 Comparison

We have computed our poverty estimation model in three different ways.

Based on MSE we can say, that the most adequate for our data is *lasso* using log(PHR) (0.02794). Then, we can conclude that for this specific data, the problem of correlation between variables is not important, as the *elastic net* is the one which has the largest MSE (4.7811).

The three computation models have similar groups of non-zero variables. Education is one of the most relevant, together with foreign relations and financial markets. However, the *elastic net* and *lasso* with log(PHR) compute a model where the explanatory variables have almost the same value each, whereas the *lasso* selects four explanatory variables as the most important.

6 Conclusions

Our research paper had the initial purpose of combining both theory and statistics to compute a simple model which can estimate the most relevant factors that influence poverty. Also, we wanted to learn how the *lasso* performs in economics studies context. This was accomplished, as almost all the variables which are relevant in the theoretical framework are also relevant in the statistical point of view. Specially education, which is one factor that we defined as important together with foreign markets. The one which surprised us is the deposit interest rate and the necessity of *liquidity* which is also important in determining poverty.

The fact that *lasso* using log(PHR) is the one with the least MSE is in accordance with most economic studies which use the *logarithmic* transformation of a variable. We thought that the correlation between variables will deteriorate the estimation, but this was not the case of data.

Lasso and *elasticity net* are computational methods that rapidly inform about the relevance of the variables you select for your model and we have also shown that it performs well in contexts were the number of variables is far greater than the number of observations (low MSE).

In conclusion, we can say that these two sparse models performed well in our data and gave us the results we expected.

References

[1] S. Edwards, *Crisis and reform in Latin America: From despair to hope.* The World Bank, 1995.

[2] G. Psacharopoulos, *Poverty and income distribution in Latin America: the story of the 1980s.* World Bank Publications, 1997, vol. 351.

[3] G. Psacharopoulos and Y. Ng., "Earnings and education in latin america: Assessing priorities for schooling investments," *World Bank, Latin America and Caribbean Region, Washington, D.C.*, vol. Working Paper 1056, 1992.

[4] J. Foreman-Peck, "A history of the world economy: International economic relations since 1850," pp. 323–354, 1995.

[5] C. E. para la América Latina y el Caribe. Cepalstat, bases de datos y publicaciones estadísticas. [Online]. Available: http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Portada.asp

[6] A. V. B. . E. Duflo, *Poor Economics: a radical rethinking of the way to fight global poverty*, 2012th ed. Random House India, 2011.

[7] *Claves de la Economía Mundial*, 11th ed. Instituto Español de Comercio Exterior, 2011.

[8] T. W. Bank. Metadata, the world bank database. [Online]. Available: http://databank.worldbank.org/data/reports.aspx?ReportId=40429&Type=Table

[9] B. Ohlin, "Interregional and international trade," *Political Science Quaterly*, vol. 49, no. 1, 1934.

[10] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity.* CRC Press, 2015.

[11] H. R. Varian, "Big data: New tricks for econometrics," *The Journal of Economic Perspectives*, vol. 28, no. 2, pp. 3–27, 2014.

[12] D. M. Chickering and D. Heckerman, "Fast learning from sparse data," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence.* Morgan Kaufmann Publishers Inc., 1999, pp. 109–115.

[13] T. Hastie and J. Qian, "Glmnet vignette," Technical report, Stanford, Tech. Rep., 2014.

Appendix

Figure 4: *lasso*[log(PHR)] coefficients plot

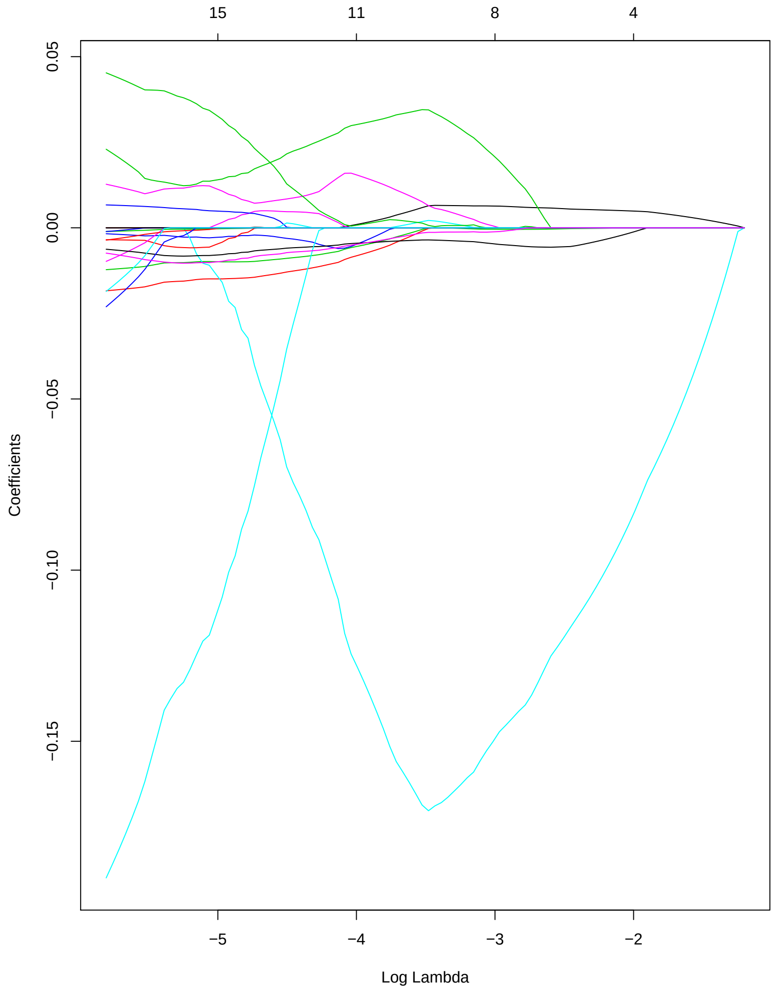


Figure 5: *lasso*[log(PHR)] MSE

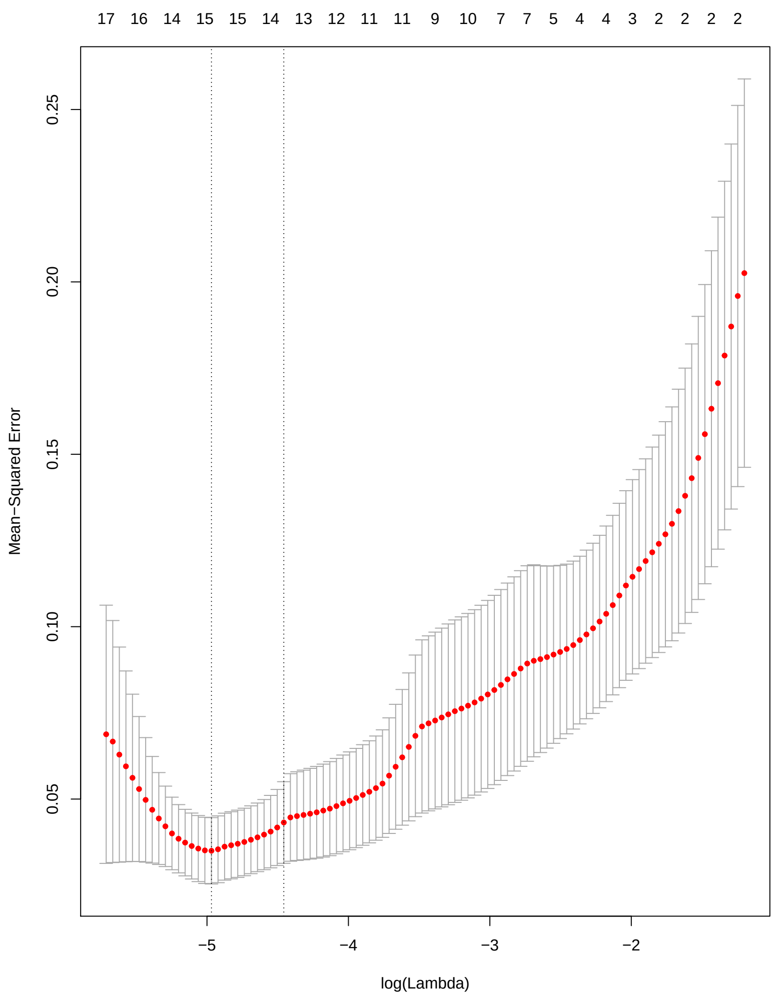


Figure 6: *elastic net* coefficients plot

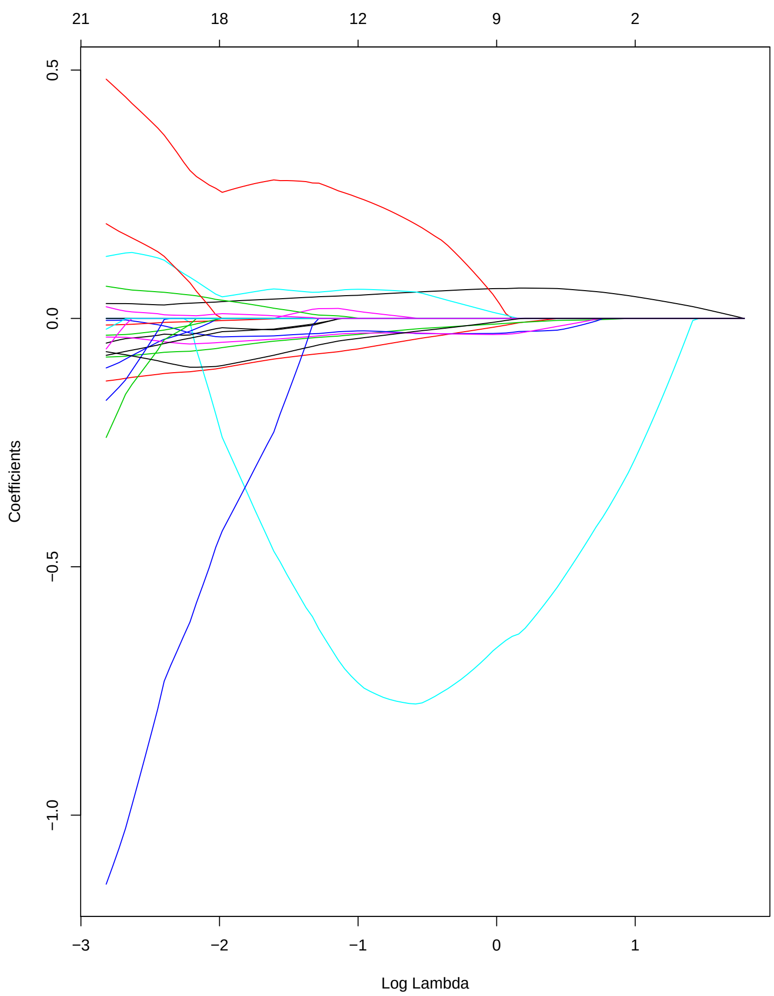
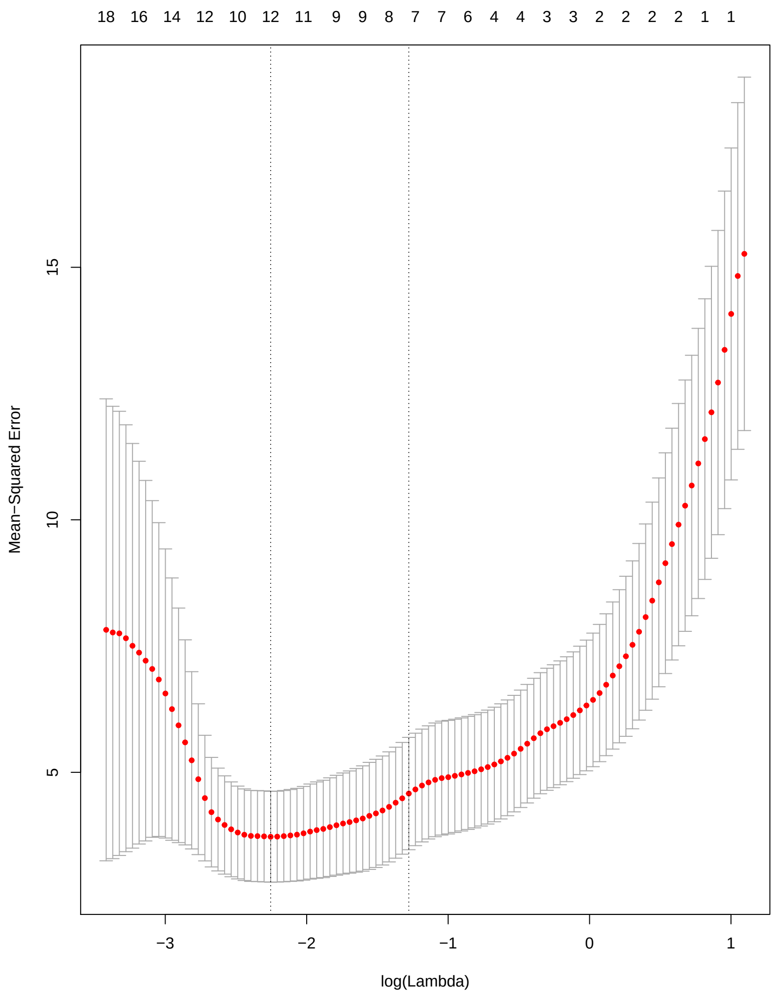


Figure 7: *elastic net* MSE



Coefficients definition table

Coefficient	Definition
PHR	Poverty headcount ratio at $3.10 a day (2011 PPP) (% of population)
FE	Fuel exports (% of merchandise exports)
FE	Fuel exports (% of merchandise exports)
FOE	Food exports (% of merchandise exports)
ME	Manufactures exports (% of merchandise exports)
SI	Adequacy of social insurance programs (% of total welfare of beneficiary households)
SPL	Adequacy of social protection and labor programs (% of total welfare of beneficiary households)
SN	Adequacy of social safety net programs (% of total welfare of beneficiary households)
EE	Adjusted savings: education expenditure [% of GNI(gross national income)]
AFR	Adolescent fertility rate (births per 1,000 women ages 15-19)
ADR	Age dependency ratio (% of working-age population)
AL	Agricultural land (% of land area)
A	Agriculture, value added (% of GDP)
EC	All education staff compensation, primary (% of total expenditure in primary public institutions)
CEF	Children in employment, female (% of female children ages 7-14)
CEM	Children in employment, male (% of male children ages 7-14)
COS	Children out of school (% of primary school age)
CSU	Cost of business start-up procedures (%GNI per capita)
CE	Cost to export $US per container
CI	Cost to import $US per container
CSS	Coverage (% All social protection and labor)
DSE	Debt service on external debt (TDS, current $US)
IN	Deposit interest rate(%)
CR	Domestic credit provided by financial sector(% of GDP)
BF	Ease of doing business index (1= most business friendly regulations)
ET	Employers total (%of employment)
EA	Employment in agriculture (% of total employment)
EI	Employment in industry (% of total employment)
ES	Employment in services (% of total employment)
GEE	Expenditure on education as % of total government expenditure (%)
FR	Fertility rate, total(births per woman)
GC	General Government final consumption expenditure (% of GDP)
CF	Gross capital formation (% of GDP)
S	Gross domestic savings (% of GDP)
EPS	Gross enrollment ratio, primary, both sexes (%)
ESS	Gross enrollment ratio, secondary, both sexes (%)
ETS	Gross enrollment ratio, tertiary, both sexes (%)