Income Micro-Determinants: A Statistical Analysis

Alessandro Tenderini and Ramon Talvi

December 21, 2022

1 Introduction

In this paper we examine the relationship between income and various individual characteristics using a micro dataset of survey data from Italy. Our aim is to predict the individual income target variable based on micro determinants for variables that can be used to identify and understand income differences among individuals.

In the first part of our project, we use penalized likelihood regression to compare the performance of lasso and adaptive lasso. This method allows us to select relevant variables and obtain interpretable results. After penalized likelihood regression models (frequenstist approach) we perform bayesian model selection and averaging. This allows us to obtain point estimates, posterior intervals, and posterior probabilities for the selected variables. In the final part of the first section, we compare the results of the frequentist and bayesian approaches in terms of variable selection and predictive performance.

In the second part of the project, thee idea is to use quantile regression to detect the variables that influence income at different levels of the distribution. The first section we use quantile regression to analyze the role of education and other relevant variables -such as demographics and job-related ones- for different quantiles of the distribution. In the second section we adopt a bayesian approach to quantile regression, also adding an adaptive lasso penalization term to limit the risk of overfiting and strengthen the process of variable selection. Here we focus on how two particular variables, wealth and job status, evolve through different quantiles and how they contrast between the extreme quantiles.

2 Related Work

One paper that has used penalized lasso regression is "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties" (Fan and Li, 2001). This paper presents a nonconcave penalized likelihood approach to variable selection, which has been shown to be more effective than traditional lasso regression in certain situations. Another paper that has used bayesian model averaging is "Bayesian Model Averaging: A Tutorial" (Hoeting, Madigan, Raftery, and Volinsky, 1999). This paper provides a comprehensive tutorial on bayesian model averaging, including its benefits and limitations. Finally, One paper that provides an introduction to bayesian regression is "Quantile Regression" (Koenker, 2010). This paper presents an overview of quantile regression and discusses its applications in various fields.

Overall, these papers demonstrate the effectiveness of penalized lasso regression, bayesian model averaging, and quantile regression in predicting continuous target variables. These approaches have been shown to be particularly useful in situations where there are many variables and a need for variable selection.

3 Dataset

The data used in this project are survey data from the Italian bank, Banca d'Italia, which can be found at https://www.bancaditalia.it/statistiche/tematiche/indagini-famiglie-imprese/bilanci-famiglie/ distribuzione-microdati/index.html. The data correspond to the year 2022 and include information on 10876 individuals, including their income and various covariates such as education level, type of work, employment history, and ability to work in a smart work environment. Before analyzing the data, we performed some preprocessing steps. Initially, we had to merge three different files according to some key values. Then, the dataset contained 16 covariates, but these were one-hot encoded using a leave-one-out method in Python, resulting in a total of 53 variables. We also imputed missing values using a logical approach and eliminated observations with negative income, resulting in a final dataset of 10577 observations. We renamed the variables according to their meaning and translated them into English. We split the dataset into a train set and a test set, with 70% of the observations being placed in the train set and 30% in the test set. This split will be used to

evaluate the predictive performance of any models developed using the data. In total, the train set consists of 7403 observations and the test set consists of 3174 observations.

4 Penalized regression

4.1 Methodology

In this section we introduce the concept of penalized regression which is a type of statistical model that introduces a penalty term on the coefficients of the regression. The purpose of this penalty term is to prevent over fitting and improve the model's generalization to new data. Such penalty term is typically chosen to be a function of the magnitude of the coefficients, and is used to constrain the coefficients so that they do not become too large. Other than preventing over fitting, we are interested in penalized models in order find best subset of important variables to include in the regression model.

In order to assess the performance of penalized regression, we compare two penalized models with the ordinary least squares (OLS) method.

4.1.1 Lasso

Lasso regression is a type of linear regression that uses L1 regularization. Such a penalty term is the sum of the absolute values of the coefficients and it is multiplied by a regularization parameter. Such regularization parameter lambda controls the strength of the penalty term and can be tuned using techniques such as cross-validated mean squared prediction error (MSPE) or BIC. The latter is better to identify truly associated variables while the former is indicated for forecasting purposes. In this project we are going to focus on the model with optimal lambda via BIC but we will also consider the other model when comparing predictive performances. The objective function for lasso regression can be written as:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda |\beta|_1$$

4.1.2 Adaptive Lasso

The Adaptive lasso regression is a variant of lasso regression that uses a modified version of the L1 penalty. Instead of penalizing the absolute values of the coefficients, the adaptive lasso penalty is a weighted sum of the absolute values of the coefficients. This is its objective function.

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^{p} \frac{|\beta_j|}{\hat{w}_j}$$

The weights are actually the inverse of estimates from another penalized model. In this project we will use ridge estimates as initial estimates. Ridge regression is another penalized model which uses L2 regularization and its objective function can be written as:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda |\beta|_2^2$$

When comparing OLS, LASSO and Adaptive LASSO we would expect large coefficients to be penalized less in the Adaptive LASSO compared to the normal LASSO. Hence, we should observe similarities between OLS and Adaptive LASSO for large estimates.

4.2 Results

Overall, we fitted 5 different model: OLS, LASSO and Adaptive LASSO, where the optimal λ for the penalized models are obtained both via BIC and cross validated MSE (for Ridge model only cross validated λ). Table 1 shows the optimal lambdas and the number of coefficients set to zero.

Model	Optimal λ	# of Zero Coefficients
LASSO-BIC	0.00672	16
LASSO-CV	0.0001482	0
Adaptive LASSO-BIC	0.00419	8
Adaptive LASSO-CV	0.001814	5

Table 1: Penalized models

4.2.1 Variable interpretation

In this part we consider the LASSO-BIC model and we interpret its results. This penalized model results in 16 variables set to zero and 37 selected variables. Some of the non-zero estimates are presented in table 2 (the other estimates are in the R code).

Variable	LASSO estimate
Educ_PS	-0.171337
Educ_HS	0.278272
Educ_PostDeg	0.374074
Educ_Bach	0.127788
CumLaude	0.089594
Y_50_200	0.138665
Qual_man_dir	0.520491
$\text{Qual}_{e}ntrepeneaur$	0.412221
Qual_Retired	0.296388

Table 2: LASSO estimate

Our research has yielded some interesting findings regarding the relationship between income and various variables.

Education appears to have a significant impact on income, with those possessing a high school diploma experiencing a 30% increase compared to those with no educational qualifications. Post degree specialization is even more lucrative, resulting in a 44% increase in income. On the other hand, obtaining a bachelor's degree only results in a 13% increase in income, which is lower than the increase seen for those with a high school diploma. Interestingly, those who quit school after primary education see a decrease in income of 18% compared to those with no education. Graduating cum laude is associated with a 9% increase in income.

In terms of employment status, our data shows that managers have the highest income, followed by entrepreneurs and retirees.

Demographic factors also play a role in income, with those coming from cities with populations of 50,000 to 200,000 experiencing a 15% increase in income compared to the base case of cities with up to 5,000 inhabitants.

4.2.2 Variable selection

In this part, we compare the behavior of three regression models: Ordinary Least Squares (OLS), LASSO, and Adaptive LASSO. Figure 1 shows the estimated coefficients for each model.

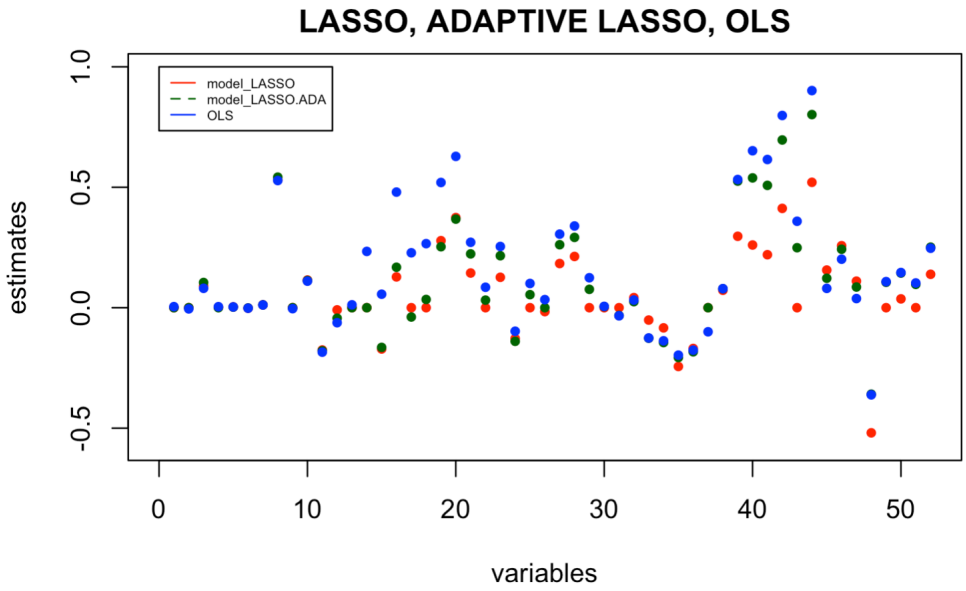

Figure 1: Plot of LASSO, adaptive LASSO and OLS estimates

As we can see, in the range of variables between 40 and 45, the Adaptive LASSO model tends to apply less penalty to large coefficients compared to the LASSO model. As a result, when the OLS estimates are large, the Adaptive LASSO estimates tend to be similar to the OLS. On the other hand, when the estimates are close to zero, both the LASSO and Adaptive LASSO models exhibit a similar behavior of shrinking the parameters to zero.

5 Bayesian Model Selection & Bayesian Model Averaging

5.1 Methods

The main idea of this section is to approach the same problem as the previous section but from a Bayesian framework. Our aim is to gather evidence on the potential variables that have a stronger predictive effect on income, analyzing the posterior probability of the top models along with the point estimates, posterior intervals, and marginal posterior probability of inclusion of the coefficients.

We define $\gamma_j = \mathrm{I}(\beta_j \neq 0)$ as the function that indicates whether parameter β_j is different from zero. We define γ=(γ_1,....,γ_d) as the set of variables that enter the model. We define β_γ as the regression coefficients given the model. We define $d_\gamma = \sum_{j=1}^{d} \gamma_j$ as the number of non-zero coefficients.

We set the prior probability of the model, $P(\gamma)$, to be $BetaBinomial(1,1)$, which implies a uniform prior on model size. Posterior probability of the model is given by:

$$P(\gamma|Y) \propto P(Y|\ \gamma).P(\gamma) \tag{1}$$

On the other hand, the integrated likelihood is given by:

$$P(Y|\gamma) = \int P(Y|\beta_\gamma)P(\beta_\gamma|\rho).P(\rho)d\beta d\rho \tag{2}$$

Note that the probability of the observed data given the model depends on the prior distribution we set on the parameters. One possible prior distribution to set on model coefficients is Zellner's prior, where $P(\beta_\gamma|\rho)$ is given by:

$$N(\beta; 0, g\rho n(X^T X)^{-1}) \tag{3}$$

The prior on the parameters depends on g. Given that we do not have a prior belief on how the dispersion of the probability distribution of the coefficients might be, we set $g = 1$, i.e., a non-informative prior.

Bayesian model selection is a method for choosing the best model from a set of candidates, using Bayesian inference to compare the posterior probabilities of different models. The first step in our analysis is to fit the model and obtain the posterior probabilities of the top performing models (we immediately check for

convergence of the gibbs sampling method in order to make sure that the number of iterations is sufficiently large to attain stability around a value).

Bayesian model averaging is a method for obtaining point estimates for the coefficients (along with posterior intervals) by combining the set of multiple models, with each model being assigned a weight based on its posterior probability. The second step is to compute the point estimates along with the posterior intervals for each coefficient and represent them in a plot. Moreover, we represent in a table the point estimates & the posterior interval for all those coefficients whose marginal posterior probability of inclusion are larger than 0.8. To complement this analysis, we represent in a table those variables whose posterior probability of the associated coefficients do not contain 0 in their posterior interval ($5\% - 95\%$).

5.2 Results

modelid	family
1,2,5,6,7,8,10,11,15,16,19,20,21,23,27,28,35,36,39,40,41,42,43,44,46,48,50,52	normal

Table 3: Top Performing Model

Analyzing the model with highest posterior probability, we detect four clusters of variables were identified as potentially having a predictive capacity for explaining individual income: educational, job-related, personal, and status as a worker. Within the educational cluster, variables such as primary school education, bachelor's degree, high school education, postgraduate degree, specialization, and graduate university education were included. In the job-related cluster, variables such as historical unemployment, minimum accepted wage, and different job fields (such as health, engineering, law, economics) were included. Industrial sector work was also included in this cluster. The personal cluster included variables such as age, whether the person is the head of their family, their sex, and their wealth. Finally, the status as a worker cluster included variables such as retirement status, employment status, worker status, autonomy, managerial or directorial status.

	estimate	X2.5.	X97.5.	margpp
Prob_knj	0.00	0.00	0.00	1.00
Age	0.01	0.01	0.01	1.00
Head_Fam	0.54	0.51	0.56	1.00
Wealth	0.11	0.11	0.12	1.00
Sex_girl	-0.19	-0.21	-0.16	1.00
Educ_Bach	0.25	0.16	0.43	1.00
Educ_HS	0.34	0.27	0.54	1.00
Educ_PostDeg	0.46	0.37	0.67	1.00
Prof_health	0.20	0.12	0.27	0.97
Prof_engin	0.21	0.14	0.28	0.97
Prof_law	0.27	0.20	0.32	1.00
Prof_econ	0.30	0.24	0.35	1.00
Employed_historical	-0.20	-0.25	-0.15	1.00
No_contr_Pensions	-0.18	-0.23	-0.13	1.00
Qual_Retired	0.53	0.45	0.61	1.00
Qual_employee	0.66	0.61	0.73	1.00
Qual_worker	0.61	0.55	0.68	1.00
Qual_entrepeneaur	0.81	0.76	0.88	1.00
Qual_aut	0.36	0.29	0.45	1.00
Qual_man_dir	0.92	0.86	0.99	1.00
Sector_ind	0.13	0.00	0.18	0.95
No_Y	-0.35	-0.41	-0.29	1.00
Y_50_200	0.17	0.12	0.27	1.00

Table 4: Point Estimates, Posterior Intervals & Marginal Posterior Probability

We want to find variables that have a strong relationship with individual income. For this, we present a table with the variables with high marginal posterior probability (above 0.9) and posterior intervals that exclude zero.

As in the previous analysis, four clusters of variables were identified as potentially having a influence in predicting individual income: educational attainment, job-related factors, personal characteristics, and employment status. The educational cluster includes variables such as bachelor's degree and postgraduate degree. The job-related cluster includes variables such as work in the health, law, engineering, or economics fields. The personal cluster includes variables such as age, family status, wealth, minimum salary accepted for work, and employment history. The employment status cluster includes variables such as retirement status, employment status, worker status, and managerial or directorial roles. These variables were found to have a strong relationship with individual income, as indicated by their high marginal posterior probability and posterior intervals that exclude zero.

The coefficients for the cluster of employment status had the greatest impact on individual income, with values greater than 0.5 in magnitude. Specifically, being an entrepreneur, manager, or employee had a more pronounced effect on income compared to being unemployed (base case), with coefficients of 0.9, 0.8, and 0.66, respectively. When exponentiated, these coefficients correspond to income increases of between 100% and 145% depending on the employment status, with the highest increase observed for entrepreneurs. These results indicate that employment status plays a significant role in predicting individual income.

6 Comparisson Lasso Regression and Bayesian Model Averaging

6.1 Variable selection

In this part we compare the variables selected by the LASSO-BIC and Adaptive LASSO-CV with the BMA. The following plot shows the estimates of the tree models and the BMA confidence intervals.

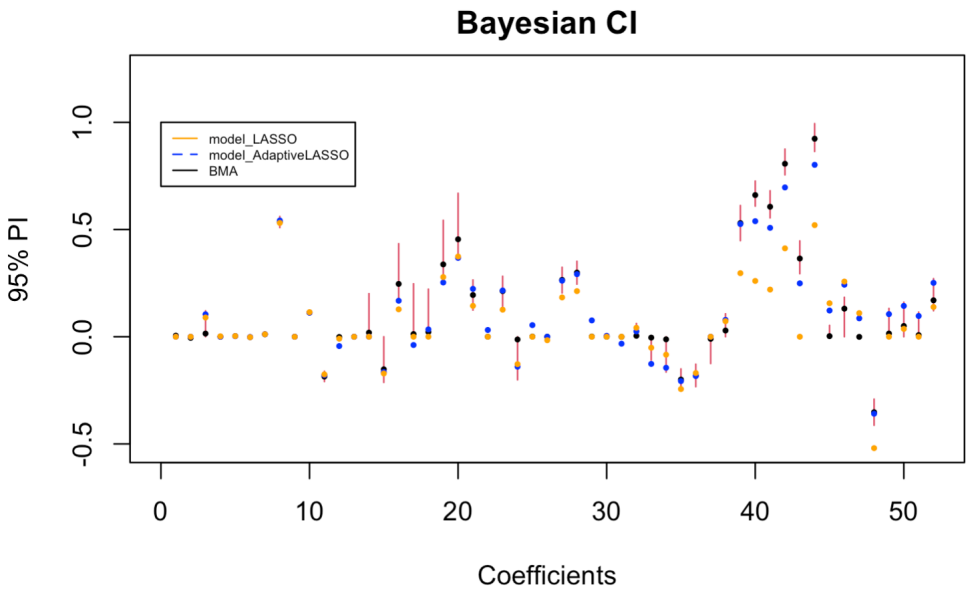

Figure 2: Plot of LASSO and adaptive LASSO and BMA confidence intervals

First of all note that the estimates for the bayesian approach and frequentest approach are relatively similar -considering confidence intervals and posterior intervals. The variables that tend to be selected and have the strongest predictive capacity for income are almost the same for both and were commented in the individual model analysis.

For our bayesian model averaging model we had fixed the prior dispersion parameter g to 1. This means our model estimation was based on a flat or uninformative prior and we know that the relationship between g and λ -penalized likelihood- is under particular circumstances inversely proportional. As we choose a diffusive prior and its incidence in posterior probability is meagre this is similar to estimating an unpenalized regression. Therefore, we see in figure 2 the lasso estimates set to exactly zero those coefficients that in bayesian model averaging are close to zero, and moreover, the large coefficients are biased towards

zero. On the other hand, the adaptive lasso controls the bias of the large coefficients towards zero with respect to penalized lasso: this is consistent with what we observe in figure 2 where the adpative lasso estimates are larger for coefficients far away from zero than in the lasso estiamte.

6.2 Predictive performances

In this part we analyze the predictive performances of bayesian model and the 5 frequentist model, OLS, LASSO-BIC, LASSO-CV, Adaptive LASSO-BIC and Adaptive LASSO-CV. Indeed, at the beginning of project we randomly split the data of 10577 rows in a train and test set (30% of the observations) and we can now use the trained models to predict the income of the test set and obtain the RMSE as a measure of predictive performance. The results are presented in the table 5.

Model	**RMSE**
LASSO-BIC	0.7625
LASSO-CV	0.7577
Adaptive LASSO-BIC	0.7568
Adaptive LASSO-CV	0.7577
BMA	0.7597
OLS	0.7577

Table 5: RMSE

Based on our analysis, the adaptive LASSO model with a λ value determined using BIC has the lowest root mean squared error (RMSE). Therefore, we actually found out that Adaptive LASSO-BIC provides good performances both in terms of prediction and also in terms of selecting the truly associated variables, as it set to zero 8 variables. Theoretically, the LASSO-CV or Adaptive LASSO-CV models should be the most accurate for prediction.

On the other hand, the LASSO-BIC results in the largest RMSE and is not recommended for prediction. This model set 16 estimates to zero, which may have resulted in some truly associated variables being mistakenly eliminated. y eliminated.

7 Quantile Regression

7.1 Methodology

In quantile regression, the goal is to estimate the relationship between the predictor variables (x) and the quantile (τ) of the response variable (y). This is done by fitting a linear model to the data, similar to traditional linear regression. However, rather than minimizing the mean squared error between the predicted and actual values of the response variable, quantile regression minimizes the quantile loss function which is the following

$$\hat{\boldsymbol{\beta}}(\tau) = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \beta x_i)$$

$$\rho_\tau(u) = u\tau - u1(u < 0)$$

The coefficients of the linear model fit by quantile regression measures the change in the value of the response variable at the th quantile as a result of a unit increase in the value of the predictor variable.

7.1.1 Lasso quantile regression

The Lasso quantile regression is a variation of the traditional quantile regression model, with the added feature of an L1 norm term. It can be implemented using the "quantreg" package in R, which uses the Frisch-Newton method to solve the corresponding optimization problem.

In our project we used the quantile lasso regression for the 0.1, 0.5, 0.9, and 0.99 quantiles of the income distribution. This can be interesting to see how the predictor variables are associated with income at the extreme points of the distribution. Indeed, variables that are selected at certain quantiles indicate that these variables may have a greater impact on the income of individuals in those parts of the distribution. This is

really interesting since the extreme quantiles of the income distribution represent specific segments of society like "rich" and "poor" people and this can reveal interesting associations and insights.

7.2 Results

We conducted quantile lasso regression on four quantiles (0.1, 0.5, 0.9, and 0.99) and found some notable associations based on selected variables and estimated coefficients.

At the 0.1 quantile, our analysis revealed that education has a significant impact on income for individuals with low income. Indeed, compared to having no education, we found that secondary school education leads to a 36% increase in income, a bachelor's degree leads to a 76% increase in income, high school education leads to a 97% increase in income, and post-degree specialization leads to a 120% increase in income. This means that the education plays a prominent role between people with low income.

In addition, our analysis showed that demographic factors, such as the size of the city where an individual comes from, also seem to affect income. Specifically, compared to coming from a city of up to 5,000 inhabitants, we found that all other larger demographic sizes are associated with approximately a 16% increase in income.

Dealing with the differences between estimates in the top 1% and bottom 10% of the distribution, we found that being an architect in the highest income quantile results in a significant 100% increase in income. However, for those in the bottom 10% of the income distribution, working as an architect actually leads to a 22% decrease in income.

Moreover, the difference in income between males and females is the largest for top quantile of 0.99. This means that in the highest level of the income distribution, we face largest income inequality by gender with a 24% decrese in income.

8 Bayesian Quantile Regression

8.1 Methods

As commented in the previous section, we are interested in analyzing which variables have a strong predictive capactiy across the different quantiles. Furthermore, the idea of this section is to compare the traditional frequentistic quantile regression results from the previous section with the ones obtained from a bayesian approach. According to Al-Hamzawi et al. (2012)[1], "the Bayesian adaptive lasso is a computationally attractive alternative to the non-Bayesian adaptive lasso, especially in high-dimensional settings, where the number of predictors exceeds the sample size (our number of predictors is fairly large). This method can be used to improve the accuracy of quantile regression models by selecting only the most important predictor variables and avoiding overfitting". This is one of the reasons we see convenient to adopt a bayesian approach to compare to the frquantist quantile regression. Additionally we a penalization term to strengthen the variable selection process.

(Adative lasso) Bayesian quantile regression is a statistical method for estimating quantiles of a continuous dependent variable while also performing variable selection using the adaptive lasso. Quantile regression models are not only more robust to outliers than ordinary least squares, but also under quantile regression one can analyze different parts of the distribution and is not limited to the mean as in traditional mean regression. In this assignment the adaptive lasso is included as a penlization term to do some additional variable selection while estimating the coefficients for different quantiles.

The model we are going to use can be expressed as:

$$Y_\tau = X\beta_\tau + \epsilon$$

where y is the continuous dependent variable, X is the design matrix, β is the vector of coefficients, and ϵ is the error term. The error term is assumed to follow an asymmetric Laplace distribution (ALD) with location parameter 0, scale parameter σ, and quantile τ. Note that unlike the Bayesian quantile regression explained in the course, we are not assuming that the prior distribution on the coefficients follows a normal

[1] Al-Hamzawi R, Yu K, Benoit DF (2012). Bayesian Adaptive Lasso Quantile Regression.(p603)

distribution but an ALD (now we have a penalty term for regularization purposes).
The priors for the model are specified as follows[2]:

$$\beta_\tau \sim \text{ALD}(\text{location} = 0, \text{scale} = \frac{\sigma}{\lambda}, \tau = 0.5)$$

$$\sigma \sim \text{InvGamma}(\text{shape} = \sigma_{\text{shape}}, \text{scale} = \sigma_{\text{scale}})$$

$$\left(\frac{\lambda}{\sigma}\right)^2 \sim \text{Gamma}(\text{shape} = \eta_{\text{shape}}, \text{scale} = \eta_{\text{scale}})$$

In this model the scale parameter λ controlls the strength of the penalty. The scale parameter σ of the error term is given an inverse-gamma prior, while the square of the ratio of λ to σ is given a gamma prior.

Al-Hamzawi et al. (2012)[3] describe how particular prior distributions can be used to develop a very efficient Gibbs sampling algorithm for the Bayesian adaptive lasso model. The choice of prior distributions in this context is motivated by the use of a location-scale mixture of normals representation of the asymmetric Laplace distribution, both in the likelihood and in the prior (due to the absence of likelihood we assume an asymmetric laplace). Recall from part 2 of the course that the Laplace quantile regression with latent variables is an alternative representation where the asymmetric laplace distribution is written as a mixture of independent exponential and normal random variables. This enables the update of the conditional probability for parameter estimation via gibbs sampling be known in closed form (linear normal gaussian). This approach is taken in the estimation of the coeffficeints for different quantiles via gibbs sampling and priors are choosen accordingly. We start from a distribution that does not belong to the exponential family and by introducing latent variables we can sample from a normal distribution.

The first thing thing we do is fit the described model for different quantiles: we set flat uninformative prior ($\sigma_{\text{shape}} = \sigma_{\text{scale}} = \eta_{\text{shape}} = \eta_{\text{scale}} = 0.01$) We do a large number of iterations and though a trace plot of different coefficients across different quantiles we make sure convergance has be attained. Afterwards, we obtain the point estimates for the relevant representative quantiles (0,1;0.5;0.9;0.99) in order to investigate if there is evidence that some variable can have a stronger predictive capacity in a given quantile. For those coefficients that we suspect that vary across the quantiles, we do a quantile plot to see how the evolve along different quantiles. The credible intervals are represented in the quantile plot with a grey shadow.

8.2 Results

We will consider a dropped variable those small coefficients whose posterior interval includes zero, given we don't have conclusive evidence to affirm they have an association with income. In simpler terms, we set to zero the coefficients that are really close to zero.

For the 0.1 percentile quantile, the following 8 variables were dropped: number of jobs, number of people in the family that are income receivers, if the person is not an italian citizen, if person assisted to a professional education high school, if person assisted to a classical type high school or other type, if person assisted to artistic high school and if person is currently unemployed.

For the median 0.5 percentile quantile, the following 3 variables were dropped: number of jobs, number of people in the family that are income receivers, if the person is not an italian citizen, and if person assisted to a classical type high school.

For the 0.9 percentile quantile, the following 4 variables were dropped: number of jobs, number of people in the family that are income receivers, if the person is not an italian citizen, and if person assisted to a classical type high school.

For the 0.99 percentile quantile, the following 6 variables were dropped: the minimum salary the person is willing to accept for a job, , number of people in the family that are income receivers, if person assisted to a classical type high school and if the person has be historically unemployed.

[2] following information extracted from the BayesQR package documentation, https://cran.r-project.org/web/packages/bayesQR/bayesQR.pdf

[3] Al-Hamzawi R, Yu K, Benoit DF (2012). "Bayesian Adaptive Lasso Quantile Regression"

Our analysis indicates there are significant differences in points estimates across quantiles, specially between the bottom 10% (0.1) and top 1% (0.99) which are the most contrasting populations. Based on the analysis of the these two populations, we are going to analyse those variable that have one of the largest absolute distance. We will focus on the work status variable employee -if person works as an employee- and on wealth. Note that as educational variables were commented in the previous section and we naturally expect them to be associated with income, we will focus on work status variable employee and wealth.

	Est_0.1	Est_0.5	Est_0.9	Est_0.99
Wealth	0.28	0.23	0.20	0.16
Qual_employee	0.62	0.15	0.08	0.01

Table 6: Quantile Point Estimate

Both wealth and employee status have strong predictive capacity for income in the bottom quantile. However, as we move up the quantiles, the predictive capacity of wealth diminishes. In other words, an increase in wealth for those in the lower quantiles will likely lead to a bigger increase in income than for the top 1%. Similarly, being an employee (with respect to base case, being unemplyed) is strongly associated with predicting income in the bottom quantile, but this effect becomes considerably weaker in the median quantile and is almost insignificant for the top 1%. There is no conclusive evidence (as indicated by the inclusion of zero in the posterior interval) that moving from unemployment to employment predicts an increase in income for the top 1%. This may be because most of the income the top 1% receives is capital based, so the role of employment on income prediction is negligible.

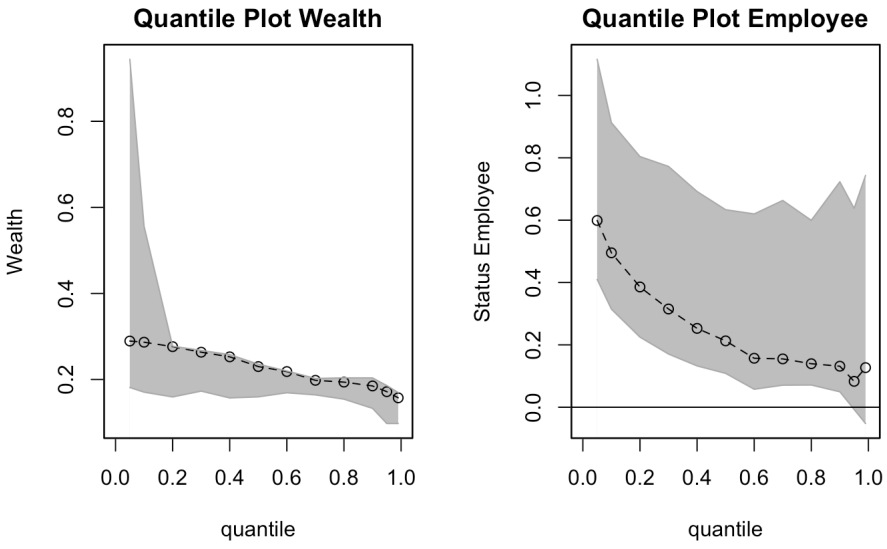

Figure 3: Quantile Plot for Wealth & Worker Status Employee

The figure above shows the evolution of wealth and employment status variables across all quantiles, represented by grey shadows indicating credible intervals. Our analysis suggests that as the quantile increases, the predictive capacity of wealth in individual income decreases. Specifically, the predictive incidence for the bottom quantile is considerable, but for the top 1%, an increase in wealth does not appear to be positively associated with income.

On the other hand, for individuals who are employed, the same relationship is observed: as the quantile increases, the predictive capacity of employment status on income decreases. However, the estimates in magnitude are much larger than those for wealth, and the contrast between extreme quantiles is more noticeable. For the bottom 1%, there is a strong positive relationship between being an employee and income, while for the top 1%, there is not conclusive evidence that there is a positive relationship (although the point estimate is positive, zero is included in the posterior interval).

9 Discussion

In this paper we examined the performance of both Bayesian and frequentist approaches for penalized regression and for quantile penalized regression. Our comparison of the LASSO and Adaptive LASSO models using different lambdas revealed that the Adaptive LASSO model with lambda selected using the BIC criterion performed the best in terms of prediction. This suggests that this model is able to identify variables that are truly associated with income. Additionally, we found strong associations between income and the highest level of education, demographic size of birthplace, sex, sector of employment, and employment status.

The bayesian model averaging suggested that there were two main clusters that have have a noteworthy association with income: those educational related and those related to the status of the worker. These last group of variables are the ones that appear to have a strong predictive capacity for income. Within this cluster, there is evidence to suggest that being an entrepreneur or a manager is the strongest association and largest coefficients in magnitude.
Comparing the penalized lassos and the bayesian model averaging we observe consistent estimates and variable selection. Also, the estimates behaves as expected: lasso coefficients are biased towards zero with respect to the bayesian model averaging given the flat prior, and the adaptive lasso mitigates this bias towards zero of the regular lasso for coefficients that are far away from zero.

From the quantile regression model we observed that education plays a prominent role across all quantiles in predicting income, the lower the quantile the stronger the association. We also discovered that the negative association between being female and income is more noteworthy in the top 1%. On the bayesian quantile regression we briefly mentioned those variables for which we had evidence that were not relevant in predicting income. We choose to focus our attention in how being an employee versus being unemployed and the level of wealth impact on income across quantiles. Both being an employee and wealth decrease their influence in predicting income as the quantile increases, being particularly low for the top 1%.

Overall, this micro dataset has a lot of potential for further analysis due to its large number of observations which have already been cleaned by the Italian Bank. The key distinctive indicator is its exhaustiveness: it measures a large set of variables that a priori we suspect could be associated to income.

If more time were available we would have considered inference on the penalized likelihood models via post selection approach or bootstrapping. We would of definitely have liked to explore the impact of different priors on the bayesian models (for bayesian model averaging set g via marginal likelihood or prior elicitation). Moreover, in the quantile section we would have structured to compare the frequentist and bayesian approach both for each variable clusters and for each quantile. Moreover, we would have presented histogram plots to observe the distribution of individual variables in each quantile (code in appendix).

A Appendices

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.

Hoeting, J. A., Madigan, D., Raftery, A. E., & Volinsky, C. T. (1999).Bayesian model averaging: A tutorial. Statistical science, 14(4), 382-401.

Koenker, R. (2010). Quantile regression. Cambridge University Press.

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

George, E. I., & McCulloch, R. E. (1997). Approaches for Bayesian model averaging. Statistical science, 12(2), 153-173.

Yang, Y., & Omelka, M. (2013). A tutorial on quantile regression. arXiv preprint arXiv:1310.5944.

Al-Hamzawi R, Yu K, Benoit DF (2012). "Bayesian Adaptive Lasso Quantile Regression."