Erasmus University Rotterdam

Erasmus School of Economics

Bachelor Thesis (Quantitative Finance)

Improving patient survival prediction with the joint-LASSO Cox model

Author
Danica van den Berg
Student ID number
470755

Supervisor
Anastasija Tetereva
Second assessor
Wendun Wang

Abstract

Using gene expression profiles to predict the survival of patients is gaining importance in the biomedical world. Good survival prediction models need to take into account the fact that genomic data is often high-dimensional and that survival data is generally right-censored. In addition, patient subgroups can be heterogeneous with regard to their underlying regression model. The problem with the latter is that estimating a separate model for each subgroup might not be possible due to small sample sizes. For this reason, a novel model, the joint-LASSO Cox, is proposed in this paper. This model allows for information sharing between subgroups along with taking into account the properties of high-dimensional data and censoring. The predictive performance of the proposed model is evaluated through simulations and application to real-life cancer patient data. Comparison of Goodman and Kruskal's γ value for nine different prediction models, including the new model, leads to the conclusion that the joint-LASSO Cox model outperforms the other models. These results indicate that gains in survival prediction of cancer patients can be obtained with the novel model.

July 2, 2020

Contents

1 Introduction

Pharmacogenetics, the study of how genes affect a person's response to drugs, is gaining importance in modern healthcare. Studying gene expression profiles can be useful to identify whether patients are likely to respond well to a medicine or to check if they are at risk for life-threatening side effects. Bhatt et al. (2010) state that recent development of microarray technology, applied to gene expression profiles, has made it possible to classify cancer patients with significantly different survival outcome by identifying potential biomarkers and gene signatures. These biomarkers and gene signatures are not only a powerful tool to distinguish patients based on their risk of death, but also to define important genes associated with risk groups (Decaux et al., 2008).

Several studies have been done to predict the survival of cancer patients from their gene expression profiles, for example Bair & Tibshirani (2004), Bullinger et al. (2004) and Lossos et al. (2004). Nguyen & Rocke (2002) were the first to construct a Cox model for survival analysis of microarray expression data. This would solve the right-censoring that is typically observed in survival data. Tibshirani (1997) extended the Cox regression model by using the LASSO technique for high-dimensional data. This model provides reasonably good predictions if we assume that all samples follow the same model. However, for cancer patients we expect different types of cancer to be non-identical with respect to the underlying regression model, possibly leading to biased results. On the other hand, constructing a different regression model for each subgroup faces the problem of loss of power due to a small sample size. To solve this problem, Dondelinger et al. (2020) came up with the joint-LASSO model which allows information sharing between subgroups and which provides group-specific estimates. In this paper, a new method called the joint-LASSO Cox, which combines the ideas of the joint-LASSO and the Cox model is proposed to predict survival of cancer patients. This model adds value to the existing literature since the model takes into account the censoring property of event data as well as the high-dimensional property of genetic data and the differences between subgroups.

This technique might be of great benefit for various medical treatment decisions, for example in the field of oncology. Over the years a lot of options for oncology treatment, like chemotherapy and radiotherapy, have been developed. In every patient that presents with cancer, doctors have to weigh the benefits of any treatment against the side-effects, which in cancer treatments often are considerable. One of the essential factors that is taken into account when making treatment decisions is the predicted survival time. Survival of patients with cancer is highly heterogeneous, from periods of a few days to more than twenty years, making it very difficult for clinicians to make a good prognosis. According to Simmons et al. (2015) inappropriate treatment near the end of life is common for many cancer patients. Therefore, the aim is to provide a better prognostic model in order to avoid unnecessary, potentially harmful treatment options.

For that purpose, different existing prognostic models will be studied, in particular the joint-LASSO model of Dondelinger et al. (2020). The underlying ideas of this model lead to the main

research question of this paper: **How can information sharing between subgroups be used to improve the survival prediction of cancer patients based on genetic information?** In order to answer this question underlying assumptions of the classical LASSO and the joint-LASSO model will be discussed. In addition, other linear and non-linear LASSO regression models are studied to see whether they can beat the prediction accuracy of the joint-LASSO. Finally, a combination of the joint-LASSO and a non-linear Cox proportional hazard model will be considered for improving the survival prediction. The predictive performance of the models is evaluated using Goodman and Kruskal's γ.

To answer the main research question, the aforementioned models are evaluated through simulation and application to cancer patient data. Characteristics of these cancer patients are obtained from cBioPortal (2019). The data set includes genomic and survival data from 1572 patients with various cancer types. These patients can be categorized into nine different subgroups based on the type of cancer they have.

The results demonstrate that the joint-LASSO Cox model has the highest predictive performance of all the compared models. The results also show robustness against changing subgroup sample size since the model has a high performance for both small and large sample sizes. This seems to indicate that combining a joint estimation technique with a penalized Cox proportional hazard model is a good step towards improvement of survival prediction models.

2 Literature review

Gene expression profiles are gaining importance both in diagnostic and prognostic predictions of various cancer types (Davicioni et al., 2010). For example, Shedden et al. (2008) showed that gene expression profiles could be used to identify potential biomarkers and gene signatures to distinguish lung cancer patients with significantly different survival outcomes. Van De Vijver et al. (2002) used complementary DNA microarrays to analyze breast-cancer tissue and to distinguish two different types of tumors with distinct patterns of gene expression. They found that the survival time of the subgroups were substantially different.

According to Algamal & Lee (2015), an important characteristic of DNA microarrays is that it is high-dimensional data. This means that the number of genes, p, exceeds the number of patients, n. This leads to statistical issues like overfitting and multicollineairty (Algamal et al., 2015). Therefore, dimensionality reduction and variable selection methods are attractive in high-dimensional data sets. One of those dimensionality reduction methods is the LASSO technique. The LASSO (Least Absolute Shrinkage and Selection Operator) regression, as formulated by Tibshirani (1996), addresses the problem of multicollinearity and increases the accuracy of linear regression models (Sermpinis et al., 2018). Xiang & Chen (2017) show that the model improves the prediction accuracy and interpretability of the statistical model by performing both variable selection and regularization.

In many cases, the total sample set can be thought of as consisting of several smaller data sets

that have similarities but that cannot be assumed to be identically distributed (Dondelinger et al., 2020). This also applies to the data set used in this paper, which consists of a large sample of cancer patients that can be divided into subsets of different cancer types. In the high-dimensional setting, estimating a different model for each subgroup is challenging due to limited sample sizes (Dondelinger et al., 2020). Huang et al. (2020) addresses this problem by proposing a linear LASSO model that integrates information on samples' tissue of origin into the feature matrix to improve the prediction of clinical drug response. They conclude that this model has a better drug response prediction as well as a more accurate identification of biomarkers of patient survival and drug sensitivity compared to the classical LASSO.

Another solution is given by Dondelinger et al. (2020). They propose the joint-LASSO model, which is a high-dimensional regression model that can be used in the group-structured setting and that allows for information sharing between groups. The model uses a penalized likelihood approach to obtain group-specific estimates. Three regularization parameters are included in the model: ρ, λ and τ. Dondelinger et al. (2020) point out that they take the same λ and ρ across subgroups. They suggest varying regularization parameters based on subgroup-specific sample size n_k for data sets with a high divergence in the number of observations in a subgroup.

In the aforementioned models a variant of the LASSO technique was integrated into linear regression models. However, the dependent variable in this paper's setting is a duration dependent variable, which means that the variable defines a length of time between two events. One of the consequences of a duration dependent variable is that censoring can occur (Kiefer, 1988). This happens when the duration process has not started before the observation period or not ended before the end of the observation period. The dependent variable, survival duration, only shows the latter of the two (right-censoring), since approximately half of the patients is not deceased yet at the end of the observation period.

Therneau & Grambsch (2000) propose to use the Cox proportional hazards model for modeling the relationship between covariates and a survival or other censored outcome. In the Cox model the non-linear hazard function, which describes the distribution of life times, forms the basis of the regression model (Aalen, 1989). The hazard function computes the risk that some event happens at a particular time. The Cox proportional hazard function relies on the assumption that the ratio of the hazards for any two individuals is constant over time (Ng'andu, 1997).

Tibshirani (1997) incorporates the LASSO technique for variable selection and shrinkage in the Cox proportional hazards model. In this way high-dimensional data can be used for the prediction of survival or other censored data. Madjar & Rahnenführer (2020) takes this model one step further by proposing a LASSO Cox regression model with a weighted version of the Cox partial likelihood that allows information sharing between subgroups. The weighted Cox model uses patients of all subgroups but assigns them individual weights based on the subgroup they belong to and how much information is shared between subgroups. In this paper, a new model will be introduced which combines the model of Madjar & Rahnenführer (2020) with the fusion approach of the joint-

LASSO. The fusion parameter τ of the joint-LASSO model will be used to determine the weights of the weighted Cox model.

3 Methodology

The predictive ability of nine different models will be compared in this paper. Six of them are variants of a linear regression model, whereas the other three are variants of a non-linear regression model. Table 3.1 provides an overview of all the models that will be discussed in more detail below.

Model	**Data**	**Objective function**	**Penalty**
OLS	full dataset	$\|\|y - X\beta\|\|_2^2$	
Pooled-LASSO	full dataset	$\|\|y - X\beta\|\|_2^2$	$\lambda\|\|\beta\|\|_1$
Subgroup-LASSO	per subgroup	$\|\|y - X\beta\|\|_2^2$	$\lambda\|\|\beta\|\|_1$
Binary-LASSO	full dataset	$\|\|y - X\beta\|\|_2^2$	$\lambda\|\|\beta\|\|_1$
Joint-LASSO	information sharing	$\sum_{k=1}^{K} \frac{1}{n_k} \|\|y_k - X_k\beta_k\|\|_2^2$	$\sum_{k=1}^{K} \left\{ \lambda \|\|\beta_k\|\|_1 + \rho \sum_{k'>k} \tau_{k,k'} \|\|\beta_k - \beta_k'\|\|_2^2 \right\}$
Joint-LASSO switching tuning parameters	information sharing	$\sum_{k=1}^{K} \frac{1}{n_k} \|\|y_k - X_k\beta_k\|\|_2^2$	$\sum_{k=1}^{K} \left\{ \lambda_k \|\|\beta_k\|\|_1 + \rho_k \sum_{k'>k} \tau_{k,k'} \|\|\beta_k - \beta_k'\|\|_2^2 \right\}$
Pooled-LASSO Cox	full dataset	$-\frac{1}{n} \sum_{i:D_i=1} (X_i\beta - log \sum_{j:T_j \geq T_i} exp(X_j\beta))$	$\lambda\|\|\beta\|\|_1$
Subgroup-LASSO Cox	per subgroup	$-\frac{1}{n} \sum_{i:D_i=1} (X_i\beta - log \sum_{j:T_j \geq T_i} exp(X_j\beta))$	$\lambda\|\|\beta\|\|_1$
Joint-LASSO Cox	information sharing	$\sum_{i=1}^{n} (\delta_i w_i \, (\beta' x_i - ln \, [\sum_{j=1}^{n} 1(t_i \leq t_j) \, w_k \, exp \, (\beta' x_j)]))$	$\lambda\|\|\beta\|\|_1$

Table 3.1: Characteristics of the different models.

3.1 Linear models

3.1.1 Standard linear model

The simplest statistical model is the **ordinary least squares (OLS) model**. The goal of this method is to find the coefficients β which solve the quadratic minimization problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \, ||y - X\beta||_2^2, \tag{1}$$

in which y represents the survival time and X a feature matrix with genomic and clinical information of all cancer patients. $||.||_p$ denotes the p-norm. This model will serve as a benchmark for the evaluation of the other models.

3.1.2 Penalized linear model

As discussed in Section 2, overfitting and multicollineairty could be a problem with OLS due to the high-dimensional data. For this reason, the LASSO penalty term is introduced. The LASSO penalty performs variable selection and produces a sparse model solution. The penalized minimization problem can be formulated as follows:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \, ||y - X\beta||_2^2 + \lambda||\beta||_1, \tag{2}$$

where λ indicates the strength of the penalization. For optimization of Equation 2 the R package glmnet (Friedman et al., 2010) is applied. In this package the parameter λ is optimized by 10-fold cross-validation.

Three variants of this model are considered. The first is the **pooled-LASSO model**. In this model a single regression is performed to solve the minimization problem of Equation 2. In this regression all cancer patients will be used to predict the β coefficients. This pooled analysis may lead to biased results when subgroups are heterogeneous with respect to the underlying regression model.

The second variant is the **subgroup-LASSO model**, in which a separate regression is performed for each subgroup. This means that Equation 2 is solved for each subgroup individually. The problem with this way of estimating is that the model only includes patients of the subgroup of interest, which could be a small sample size. When the sample size n becomes relatively small compared to the number of features p, this can cause loss of power.

The third model is the **binary-LASSO model** formulated by Huang et al. (2020). Instead of estimating a different regression for every subgroup, they incorporate the subgroup information into a feature matrix. For each subgroup, a new binary feature is introduced. The feature matrix (A) then becomes an n x K matrix, where n is the total sample size and K the number of subgroups, with

$$A_{i,j} = \begin{cases} 1 & \text{if patient } i \text{ is in subgroup } j \\ 0 & \text{if patient } i \text{ is not in subgroup } j \end{cases}$$

The feature matrix A is merged with the original feature matrix of genetic and clinical variables. Then, the classical LASSO algorithm, as formulated in Equation 2, will be applied to the extended data set in order to predict the survival.

3.1.3 Joint-LASSO model

The **joint-LASSO model** introduced by Dondelinger et al. (2020) tackles the task of providing group-specific estimates with global sparsity that allows information to be shared across different cancer type samples. The joint-LASSO criterion consists of three parts: the least squares objec-

tive function, the LASSO penalty factor and a fusion constraint. The fusion constraint takes the similarities between subgroups into account. This leads to the following criterion:

$$\hat{B} = \underset{B}{\operatorname{argmin}} \sum_{k=1}^{K} \Big\{ \frac{1}{n_k} ||y_k - X_k \beta_k||_2^2 + \lambda \, ||\beta_k||_1 + \rho \sum_{k'>k} \tau_{k,k'} ||\beta_k - \beta_k'||_2^2 \Big\}, \tag{3}$$

in which $k \in \{1, ..., K\}$ indicates the specific cancer type of the patient. Every cancer type k has the same number of independent variables p, but a different sample size n_k. For every cancer type k, y_k is an n_k x 1 vector with the survival times in months, X_K is an n_K x p feature matrix and β_k is a p x 1 vector with regression coefficients. Both X_k and y_k have been standardized per subgroup. The total matrix of coefficient estimates for all cancer types (p x k) looks as follows: $B = [\beta_1 ... \beta_k]$. λ, ρ and τ are tuning parameters and, in this case, λ and ρ are the same for every subgroup.

The parameter τ, which represents the amount of fusion between two cancer types, will be defined as follows:

$$\tau_{k,k'} = 1 - d(k, k')/d_{max}$$

$$d(k, k') = ||\mu_k - \mu_{k'}||_2,$$

where d_{max} is the maximum distance between any pair of groups (k, k').

As described in Dondelinger et al. (2020), to optimize Equation 3 the problem is transformed into a classical LASSO problem and then the R package glmnet (Friedman et al., 2010) is used to find $\hat{B}$. The transformation consists of two steps. First X_{diag}, b_{flat} and y_{flat} are defined as

$$X_{diag} = \begin{bmatrix} X_1 & & \\ & \ddots & \\ & & X_K \end{bmatrix} \qquad b_{flat} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix} \qquad y_{flat} = \begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix}$$

Next, the fusion constraint is included by defining X_{new} and y_{new} as

$$X_{new} = \begin{bmatrix} X_{diag} \\ P \end{bmatrix} \qquad y_{new} = \begin{bmatrix} y_{flat} \\ \vec{0} \end{bmatrix}$$

where $\vec{0}$ is a vector of zeros and P is a matrix which represents the fusion constraints. The details on P can be found in the paper of Dondelinger et al. (2020). With X_{new} and y_{new}, the transformed problem can be represented as

$$\hat{b_{flat}} = \underset{b_{flat}}{\operatorname{argmin}} \, ||y_{new} - X_{new} \, b_{flat}||_2^2 + \lambda ||b_{flat}||_1, \tag{4}$$

which is a classical LASSO problem. For the transformation of the problem, the R package fuser (Dondelinger & Wilkinson, 2018) is used.

In order to determine the values of the tuning parameters λ and ρ, grid search is used in combination with cross-validation. Since the subgroups in most data sets differ substantially in sample size, improved prediction accuracy might be achieved by allowing tuning parameters to depend on n_k.

This idea is incorporated in the final linear model, the **joint-LASSO with switching tuning parameters**. For this model, the tuning parameters ρ and λ are allowed to be different for every cancer type k. Consequently, the parameter tuning problem becomes k-dimensional with subgroup-specific tuning parameters. One problem that arises is that grid search for a large number of parameters is computationally intensive. To avoid the exhaustive enumeration of all possible combinations, the tuning parameters are defined as follows:

$$\lambda_k = \frac{1}{n_k}\lambda \qquad\qquad \rho_k = \frac{1}{n_k}\rho$$

In this way, a univariate problem for λ and ρ remains, instead of the initial k-dimensional parameter selection problem.

3.2 Non-linear models

As discussed in Section 2, one problem with the linear prediction models is that none of them takes right-censoring into account. Therefore, the linear models are also compared to three non-linear models, including the new joint-LASSO Cox model. To fit the non-linear models and obtain the regression coefficients, the R package WeightedCoxRegression (Madjar, 2020) is used.

3.2.1 Penalized Cox model

When studying the dependence of survival time T on a set of features, Cox's proportional hazard model (Cox, 1972) assumes that

$$\lambda(t|X_i) = \lambda_0(t)\ exp\,(X_i\beta), \tag{5}$$

where $\lambda(t|X_i)$ is the hazard at time t for patient i, $\lambda_0(t)$ is an arbitrary baseline hazard function and $\beta = (\beta_1, ..., \beta_p)^T$ is an unknown vector of regression coefficients. The likelihood that patient i will die at time T_i can be written as:

$$L_i(\beta) = \frac{\lambda(T_i|X_i)}{\sum_{j:T_j \geq T_i} \lambda(T_i|X_j)} = \frac{exp(X_i\beta)}{\sum_{j:T_j \geq T_i} exp(X_j\beta)} \tag{6}$$

The joint probability of all realized events (deaths), where the occurrence of death is indicated by $D_i = 1$, can then be formulated as follows:

$$L(\beta) = \prod_{i:D_i=1} L_i(\beta) \tag{7}$$

Consequently, the log partial likelihood becomes:

$$l(\beta) = \sum_{i:D_i=1} (X_i\beta - log \sum_{j:T_j \geq T_i} exp(X_j\beta)) \tag{8}$$

To select the important variables under the proportional hazards model, Tibshirani (1997) suggests to incorporate the LASSO penalty. This leads to the following penalized log partial likelihood function:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \Big\{ -\frac{1}{n} l(\beta) + \lambda \, ||\beta||_1 \Big\} \tag{9}$$

In the classical LASSO Cox model, all patients contribute equally to the estimation of β. This **pooled-LASSO Cox model** is the first non-linear model that will be considered in this paper. However, similarly as with the linear models, when there are heterogeneous subgroups included in the data set, it might be better to fit a separate Cox model for each subgroup. Therefore, in addition to the pooled Cox model, a **subgroup-LASSO Cox model** is constructed. This model solves Equation 9 for each subgroup individually.

3.2.2 Joint-LASSO Cox model

Madjar & Rahnenführer (2020) constructed a model in which all patients contribute to the estimation of the regression coefficients. However, each patient contributes with an individual weight $w_i \in [0, 1]$ to take into account the heterogeneity of the data set. The higher the weight, the more influential the patient is for the estimation of β. Madjar & Rahnenführer (2020) used three different methods for the estimation of the weights: multinominal logistic regression with lasso penalty, ridge penalty and random forest. The log partial likelihood of the weighted version of the LASSO Cox model is defined as follows:

$$l(\beta) = \sum_{i=1}^{n} (\delta_i w_i \, (\beta' x_i - ln \, [\sum_{j=1}^{n} 1(t_i \leq t_j) \, w_j \, exp \, (\beta' x_j)])), \tag{10}$$

in which δ_i is the event indicator and t_i the observed time until an event or censoring.

This weighted version of the LASSO Cox model is extended to the novel **joint-LASSO Cox model**. In this new model the partial likelihood of Equation 10 is used in combination with the fusion parameter τ of the linear joint-LASSO model. In this way, there is more information sharing between subgroups that show similarities in features. The weights are not estimated with the multinominal logistic regression as done in the paper of Madjar & Rahnenführer (2020), but they are defined as fixed values based on the fusion parameter τ (see Section 3.1.3). Specifically, this means that when the survival time is predicted for all patients in subgroup s, the weight w_i is defined as

$$w_i = \begin{cases} 1 & \text{if patient } i \text{ is in subgroup } s \\ \tau_{s,k} & \text{if patient } i \text{ is in subgroup } k \neq s \end{cases}$$

For comparison, in the pooled-LASSO Cox model w_i is 1 for every patient and in the subgroup-LASSO Cox model w_i is 1 for the patients in subgroup s and 0 for the other patients.

Just like in the linear joint-LASSO model, adjustments could be made for diversity in subgroup-

specific sample size. If the joint-LASSO with switching tuning parameters turns out to perform significantly better than the standard joint-LASSO, then the tuning parameters of the joint-LASSO Cox model will also be adjusted. In that case, λ_k will be defined in the same way as in Section 3.1.3.

3.3 Performance measure

To compare the models Goodman and Kruskal's γ (Goodman & Kruskal, 1954) is used. The γ coefficient is a measure of rank correlation between two variables X and Y. In this setting, X is the prediction of the survival time and Y is the observed survival time. Let us consider two pairs of observations (X_i, Y_i) and (X_j, Y_j), where $i \neq j$. Goodman & Kruskal (1954) define two probabilities: the concordance probability (Π_c) and the discordance probability (Π_d).

$$\Pi_c = Pr\{(X_i < X_j) \wedge (Y_i < Y_j)\} + Pr\{(X_i > X_j) \wedge (Y_i > Y_j)\}$$

$$\Pi_d = Pr\{(X_i < X_j) \wedge (Y_i > Y_j)\} + Pr\{(X_i > X_j) \wedge (Y_i < Y_j)\}$$

The theoretical γ is defined using the concordance and discordance probability:

$$\gamma = \frac{\Pi_c - \Pi_d}{\Pi_c + \Pi_d}$$

Consequently, γ can be estimated as follows:

$$\hat{\gamma} = \frac{N_c - N_d}{N_c + N_d},$$

where N_c and N_d are the number of concordant and discordant pairs, respectively. The values for N_c and N_d can be obtained from an ordered contingency table. The first row of this table represents the group that has the smallest values for X and the first column represents the group that has the smallest values for Y. Each cell of the table then shows the number of pairs whose values of X and Y correspond to the row and column of that particular cell.

γ takes a value between -1 and 1, which indicates to what extent the ordering of X matches the ordering of Y. If γ is 1, this means that the patients with the longest predicted survival time also have the longest observed survival time. If γ is -1, the patients of X and Y are exactly in opposite order. Therefore, the higher the γ coefficient, the better the predictive ability of the model. Hence, to compare the prediction performance of the different models, the mean γ value of all testing sets is computed per model.

Goodman & Kruskal (1954) assume that the contingency table has a multinomial distribution of counts. They show that under the assumption of a multinomial distribution, $\hat{\gamma}$ has an asymptotic normal distribution. Chen & Kianifard (1999) show that the ratio of $\hat{\gamma}$ to its standard error can be used to compute the z-score for the null hypothesis $\gamma = 0$. Consequently, the test statistic z can be compared to the percentiles of the standard normal distribution to find the p-value.

Since $\hat{\gamma}$ is asymptotically normally distributed, a Student's t-test can be used to determine if two models have statistically significant different mean values for γ. To test this hypothesis simultaneously for the nine models, an ANOVA test is used. ANOVA is a multiple testing method, which means that multiple hypothesis are tested at the same time. Each hypothesis has its own type I error rate. When the number of hypotheses increases, the overall type I error rate also increases since the individual error rates accumulate. In order to keep the overall significance level below 5%, the p-values require adjustment. Le Clerg (1957) recommends to use Duncan's Multiple Range Test as a correction method for the p-values since this test also assures a good type II error protection.

4 Data

To evaluate the models both simulated data as well as real data obtained from cancer patients are used.

4.1 Simulated cancer patient data

The data set is simulated as described by Madjar & Rahnenführer (2020). Four subgroups are simulated from two differently distributed groups denoted by index $g = 1, 2$. Each group includes two subgroups, hence the data set consists of the subgroups 1A, 1B, 2A and 2B. The data within one group are simulated with the same parameters. For each subgroup, features are sampled from the multivariate normal distribution $\mathcal{N}(\mu_g, \Sigma)$, where $\Sigma = I_{p\times p}$ and μ_g is based on real gene expression data. Hebenstreit & Teichmann (2011) state that log transformations are generally applied to gene expression data since expression data are strongly skewed. After transformation to a $\log_2$ scale, real gene expression data generally vary between 4 and 12. In this simulation setting, the elements of μ_g are therefore defined as a linear function of a constant term equal to 4 and the parameter $\epsilon \in [0, 1]$. ϵ determines the extent to which there is similarity between the two groups. In this paper, ϵ will be equal to 0.5, so that there are both similarities and differences between the groups. Madjar & Rahnenführer (2020) have also analyzed different covariance structures for Σ, but they did not find significantly different results.

Bender et al. (2005) provide techniques to generate survival times for simulation studies regarding Cox proportional hazards models. They show that appropriate survival times can be generated with the Weibull distribution. Therefore, the survival data (T_g) in this paper is also simulated from a Weibull distribution, with scale parameter λ_g and shape parameter κ_g. To obtain simulation settings that match real cancer patient data, the parameters are determined based on two independent lung cancer cohorts (GSE37745 and GSE50081). These parameter values are obtained from the supplementary materials of Madjar & Rahnenführer (2020). The censoring times (C_g) are simulated from a Weibull distribution with the same parameters as the event times. Hence, for both groups, approximately 50% of the data is censored. This is similar to the real data setting, described in Section 4.2, where also approximately half of the data is censored. Whether data is censored data

or not is indicated by the variable

$$\delta_g = \begin{cases} 0 & \text{if } T_g \geq C_g \\ 1 & \text{otherwise.} \end{cases}$$

and the time until death or censoring is defined as $t_g = \min(T_g, C_g)$.

4.2 Real cancer patient data

The real data used in this paper is obtained from cBioPortal (2019). The data set includes genomic and survival data from 1661 patients with various cancer types. These patient can be categorized into nine different subgroups based on the type of cancer they have (see Table 4.1). One patient is excluded from the data set since this is the only patient with non-melanoma skin cancer. Also, the patients of which the type of cancer is unknown are removed from the data set. This reduces the sample set to 1572 patients. For every patient the data set indicates which genes are mutated. A selection of 100 possible mutated genes have been selected based on research done by Cheng et al. (2015). According to Cheng et al. (2015), the genes that were chosen play a critical role in the development and behavior of tumors.

In addition to the cancer type and mutated genes, the variables included in the data set are: mutation count, survival status, sex, drug type used for treatment, survival in months, the Tumor Mutation Burden (TMB) score and the tumor purity. The TMB score represents the total number of mutations per tumor genomic region analyzed and the tumor purity is the proportion of cancer cells in the tumor tissue. A summary of these variables is provided in Table 4.2.

Bladder cancer	215
Breast cancer	44
Colorectal cancer	110
Esophagogastric cancer	126
Brain cancer	117
Head and neck cancer	139
Skin cancer	320
Lung cancer	350
Kidney cancer	151

Table 4.1: Number of observations per cancer type.

Variable	**Values**	
Sex	male	989
	female	583
Mutation count	min.	1
	max.	212
	mean	12.6
	st.dev.	19.0
Drug type used for treatment	PD-1/PDL-1	1230
	Combo	244
	CTLA4	98
TMB score	min.	0
	max.	209.5
	mean	12.0
	st.dev.	18.8
Tumor purity	min.	0
	max.	100
	mean	45.9
	st.dev.	21.5
Survival status	deceased	796
	living	776
Survival in months	min.	0
	max.	80
	mean	14.5
	st.dev.	12.0

Table 4.2: Summary of variables.

5 Results

5.1 Simulation

First of all, the joint-LASSO model is compared to the joint-LASSO model with switching tuning parameters in order to see whether the tuning parameters of the joint-LASSO Cox model should also be adjusted. The expectation is that for widely divergent subgroup-specific sample sizes, the adjustment of the tuning parameters has a positive effect on the predictive performance. Therefore, the following sample sizes are chosen: $n_{1A} = 20$, $n_{1B} = 50$, $n_{2A} = 100$ and $n_{2B} = 500$. Figure 5.1 shows the average γ over all subgroups of the two models. The average γ values of the joint-LASSO model and the joint-LASSO model with switching tuning parameters are equal to 0.444 and 0.176, respectively. The t-statistic is 7.281, with a p-value of 0.000. Thus, the joint-LASSO performs significantly better than the joint-LASSO model with switching tuning parameters. Hence, the adjustment of the tuning parameters does not improve the predictive ability and therefore it will not be applied to the joint-LASSO Cox model.

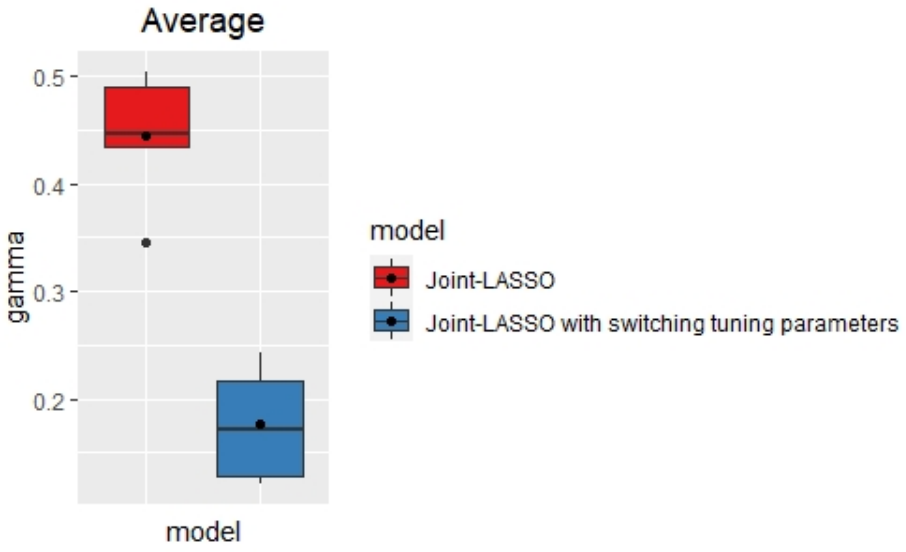

Figure 5.1: Values for average γ over all subgroups for sample sizes $n_{1A} = 20$, $n_{1B} = 50$, $n_{2A} = 100$ and $n_{2B} = 500$ and the number of features equal to 100.

Figure 5.2 shows the performance of the nine different models per subgroup and the average over all subgroups. In this setting, the subgroup sample size n and number of features p are fixed at 100.

One conclusion that can be drawn from Figure 5.2 is that the non-linear models have better predictive abilities than the linear models. This is confirmed in Table 5.1, which shows a significant difference in mean γ value between the non-linear and linear models. This supports the hypothesis described in Section 2 that using non-linear models to analyze survival data should be preferred to using linear models due to censoring.

Among the three non-linear models, the joint-LASSO Cox model has the best predictive abilities, both when looking at the mean γ value of the subgroups as well as for each subgroup individually. Table 5.1 shows that the joint-LASSO Cox model on average beats every other model significantly. Tables 7.1 to 7.4 in the appendix show that for subgroup 2A and 2B the joint-LASSO Cox model also beats every other model. However, for subgroup 1A and 1B, the difference with the subgroup-LASSO Cox is not significant. Since the sample size per group is relatively large (n=100), estimating a LASSO-Cox model for each individual subgroup is likely to obtain relatively good results. This could explain the insignificance between the joint-LASSO Cox and the subgroup-LASSO Cox model.

Another conclusion that can be drawn from Figure 5.2 and Table 5.1 is that for the prediction of survival data the linear joint-LASSO model is not a useful model since on average it does not perform significantly better than OLS. This does not mean that information sharing between subgroups is not useful, but merely that information sharing in this linear form is not suitable for survival data.

All the other models, on the other hand, reject the null hypothesis of equal means with OLS substantially. From this observation, it can be concluded that penalization, in particular LASSO penalization, significantly improves prediction.

On average there is no significant difference between the linear pooled-LASSO, subgroup-LASSO and binary-LASSO. However, Figure 5.2 shows that for an individual subgroup there can be differences in predictive ability between these three models. For example, in subgroup 1B, the subgroup-LASSO performs significantly worse compared to the pooled- or binary-LASSO.

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.288									
Pooled-LASSO	0.419	0.130 (0.001*)								
Subgroup-LASSO	0.434	0.146 (0.000*)	0.015 (0.669)							
Binary-LASSO	0.435	0.147 (0.000*)	0.017 (0.663)	0.001 (0.971)						
Joint-LASSO	0.252	-0.036 (0.316)	-0.166 (0.000*)	-0.182 (0.000*)	-0.183 (0.000*)					
Joint-LASSO switching parameters	0.241	-0.047 (0.214)	-0.178 (0.000*)	-0.193 (0.000*)	-0.194 (0.000*)	-0.012 (0.747)				
Pooled-LASSO Cox	0.569	0.281 (0.000*)	0.150 (0.000*)	0.135 (0.001*)	0.134 (0.001*)	0.317 (0.000*)	0.328 (0.000*)			
Subgroup-LASSO Cox	0.633	0.345 (0.000*)	0.214 (0.000*)	0.199 (0.000*)	0.198 (0.000*)	0.381 (0.000*)	0.392 (0.000*)	0.064 (0.078)		
Joint-LASSO Cox	0.717	0.429 (0.000*)	0.299 (0.000*)	0.283 (0.000*)	0.282 (0.000*)	0.465 (0.000*)	0.477 (0.000*)	0.148 (0.000*)	0.084 (0.022*)	

Table 5.1: Mean γ and differences in mean γ of the average over all subgroups.
* indicates a significant differences in mean for a 5% significance level

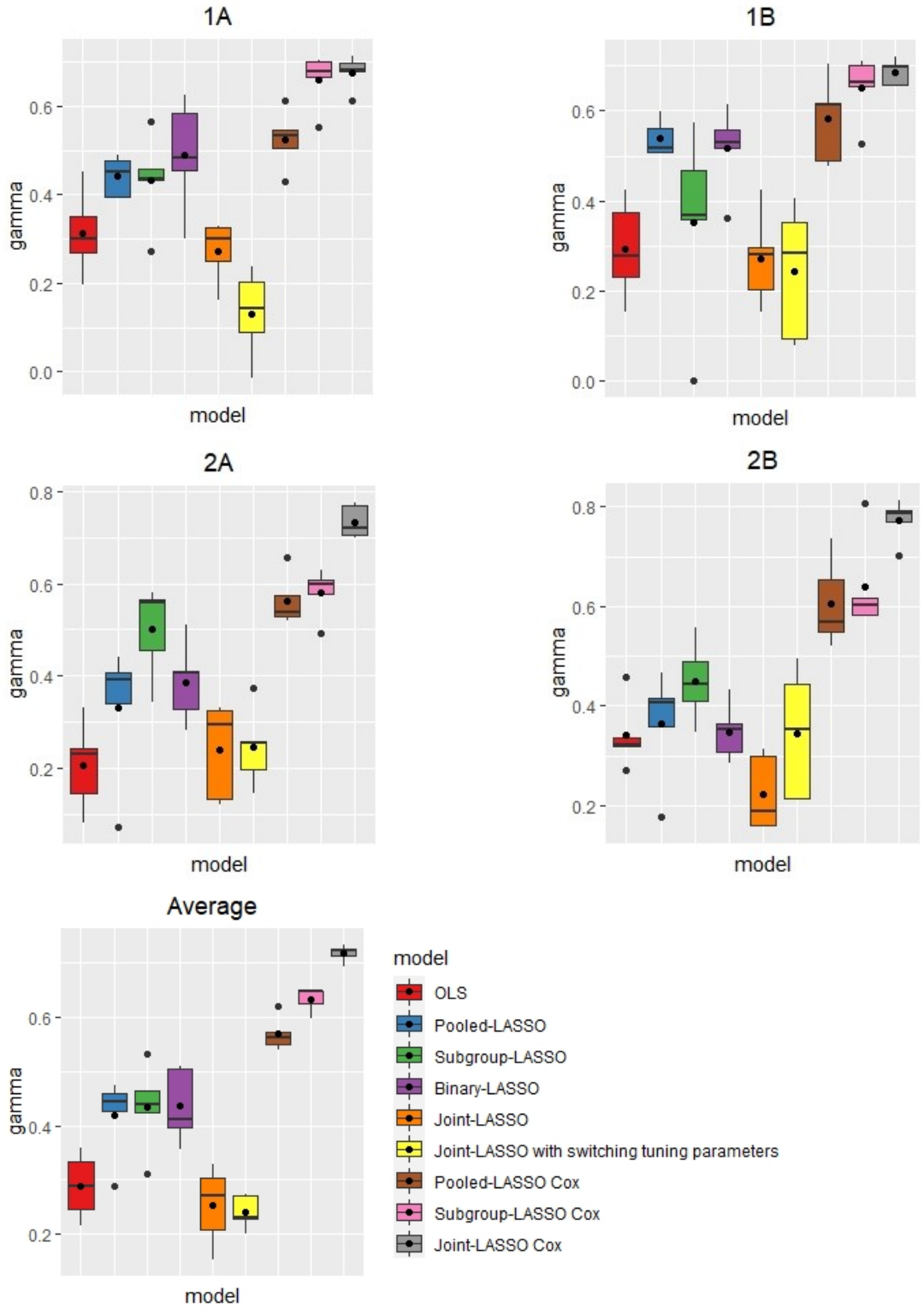

Figure 5.2: Boxplots of γ for the prediction of each subgroup and the average of the subgroups with both the subgroup sample size and number of features equal to 100.

Since the predictive ability can vary substantially with a different sample size, the average γ over all subgroups is not only obtained for n equal to 100. Figure 5.3 shows the average γ over all subgroups for varying subgroup sample sizes. It proves that the non-linear models do not only have the best predictive ability for n equal to 100, but for a large range of sample sizes.

The three non-linear models show an expected pattern. For a small sample size, the pooled method performs best, because it can use observations from all subgroups. The subgroup method performs the least, since it only has 20 observations to estimate the model. The joint-LASSO is in between the two aforementioned. When the sample size increases, the predictive ability of the subgroup method increases and exceeds the predictive ability of the pooled method. This is caused by the fact that the subgroup method can obtain different estimates for each subgroup which are more accurate when the underlying regression models of the different subgroups are heterogeneous. The joint-LASSO Cox beats the predictions of the subgroup method and pooled method for intermediate sample sizes (n between 50 and 500). An explanation for this finding is that the joint-LASSO can use the characteristics of both the pooled and subgroup method. It estimates different regression coefficients for every subgroup like in the subgroup method, but in addition it can use information from all observations like the pooled method.

Moreover, Figure 5.3 confirms that the linear joint-LASSO model is not an appropriate model for survival prediction since it performs significantly worse compared to the linear subgroup and pooled method for n between 20 and 500.

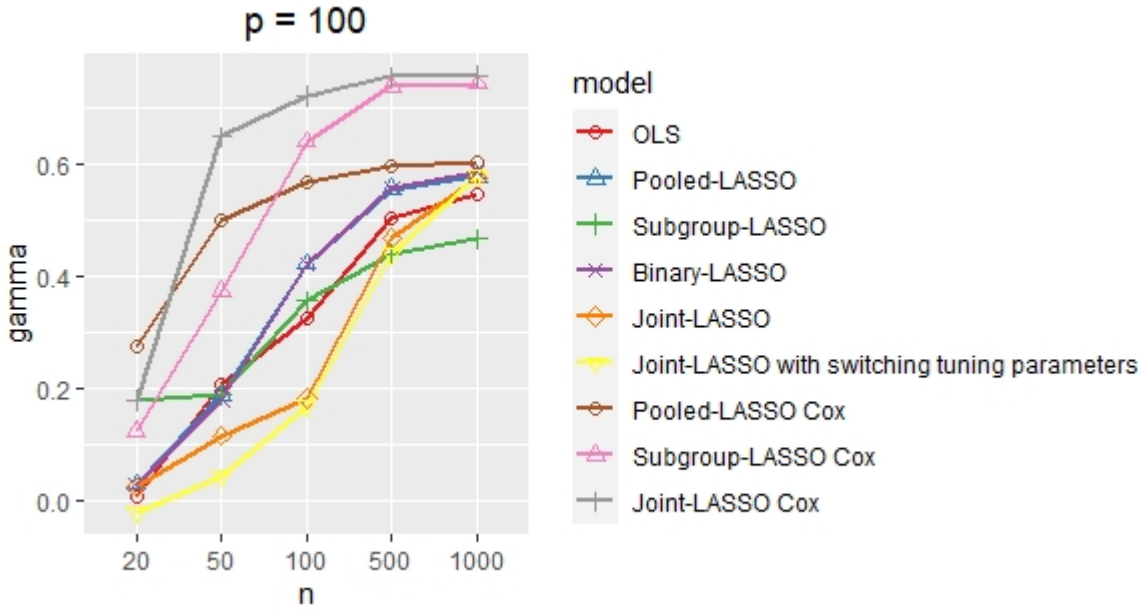

Figure 5.3: Values for average γ over all subgroups for different sample sizes per subgroup(n) with the number of features equal to 100.

The predictive ability could also be different for a varying number of features. For that reason, in

Figure 5.4, the predictive abilities are also compared for p equal to 12. In comparison to Figure 5.3, the linear models perform substantially better for a sample size between 50 and 500. This reduces the gap between the non-linear models and the linear models. Although the performance of the linear models is better for a smaller number of features, the joint-LASSO Cox and the subgroup-LASSO Cox models still have substantially better γ values for n above 50. For p equal to 12, the joint-LASSO Cox model even has the best performance for very small sample sizes.

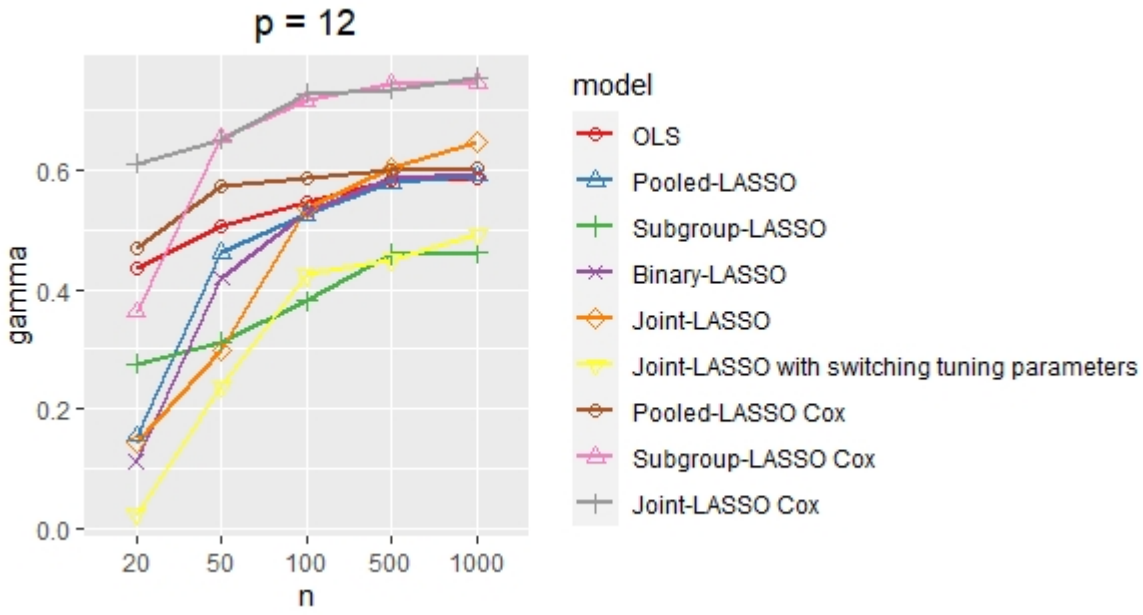

Figure 5.4: Values for average γ over all subgroups for different sample sizes per subgroup(n) with the number of features equal to 12.

5.2 Real data

5.2.1 Comparison models

Figure 5.5 shows the performance of the different models on average over all the different cancer types. For this real data set the differences in prediction performance are a lot smaller than in the simulation case. Table 5.2 shows that only a few differences in mean are significant. However, even though the differences are small, the joint-LASSO Cox model still seems like the best model to use for the survival prediction since Table 5.2 shows that the joint-LASSO Cox model has the highest mean γ value. A significant difference with the mean γ value of the joint-LASSO Cox method is observed for OLS, pooled-LASSO Cox and subgroup-LASSO Cox. Thus, contrary to the simulation case, the joint-LASSO Cox is not significantly better than the penalized linear models.

In addition, not all non-linear models perform better than the linear models as in the simulation case. The pooled-LASSO Cox model and the subgroup-LASSO Cox model have a γ value that is not even significantly better than OLS.

The results for each individual cancer type are included in Figure 7.1 and Tables 7.5 to 7.13 in the appendix. Similar to the average over all subgroups, the individual results also show only very small differences between the different prediction models.

As discussed in Section 2, the Cox proportional hazard function relies on the assumption that the ratio of the hazards for any two individuals is constant over time. It could be possible that the real data set lacks the hazard proportionality, which would explain why the non-linear Cox models do not show significantly better results than the linear models.

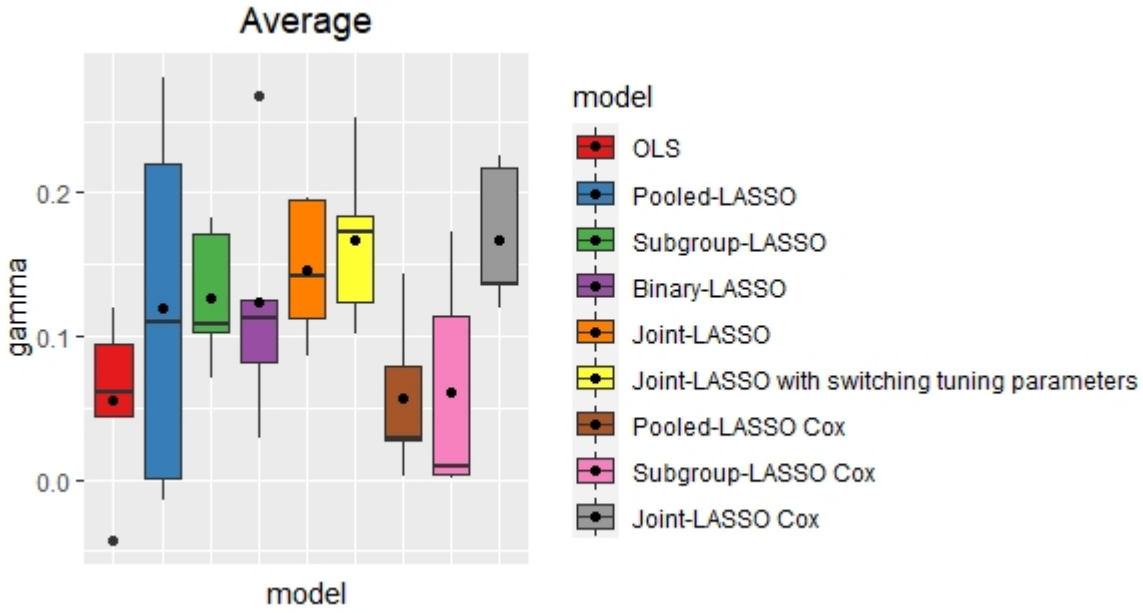

Figure 5.5: Boxplot of γ for the average prediction over all cancer types.

	Mean	Differences in mean								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.055									
Pooled-LASSO	0.119	0.065 (0.215)								
Subgroup-LASSO	0.127	0.072 (0.182)	0.008 (0.877)							
Binary-LASSO	0.123	0.068 (0.201)	0.004 (0.937)	-0.004 (0.931)						
Joint-LASSO	0.146	0.091 (0.096)	0.027 (0.604)	0.019 (0.683)	0.023 (0.642)					
Joint-LASSO switching parameters	0.167	0.112 (0.045*)	0.048 (0.378)	0.040 (0.443)	0.044 (0.409)	0.021 (0.680)				
Pooled-LASSO Cox	0.056	0.001 (0.977)	-0.063 (0.209)	-0.071 (0.183)	-0.067 (0.199)	-0.090 (0.097)	-0.111 (0.046*)			
Subgroup-LASSO Cox	0.060	0.005 (0.916)	-0.059 (0.211)	-0.067 (0.198)	-0.063 (0.211)	-0.086 (0.106)	-0.107 (0.052)	0.004 (0.933)		
Joint-LASSO Cox	0.168	0.112 (0.044*)	0.047 (0.374)	0.040 (0.430)	0.044 (0.401)	0.020 (0.664)	0.001 (0.995)	0.111 (0.045*)	0.107 (0.050*)	

Table 5.2: Mean γ and differences in mean γ of the average over all subgroups.
* indicates a significant differences in mean for a 5% significance level

5.2.2 Estimation regression coefficients

Since the joint-LASSO Cox method has the highest mean γ value, this model is used to give an answer to the question: "Which variables are important for the prediction of survival duration?" Figure 5.6 shows the regression coefficients with a magnitude above 0.05 from three cancer groups. Some variables are influential for the survival prediction of all three cancer types, like $Drug.Type$, $EPHA7$, $PAK5$ and $TP53$. $Drug.Type$ has a negative coefficient, which indicates that patients receiving the drug PD-1/PDL-1 on average have a higher survival time compared to patients receiving CTLA4. This negative relation between $Drug.Type$ and survival time showed up in every cancer type (see Table 7.14 in the appendix for the complete coefficient list of every type). $EPHA7$ and $PAK5$ also have negative coefficients for every cancer patient subgroup, indicating that patients with mutations in these genes, have a shorter expected life duration than patients without mutations in these genes.

$TP53$, on the other hand, has a positive coefficient. However, this cannot be interpreted as having a higher life expectancy when you have the TP53 gene mutation. Mutations in the TP53 gene are frequent in most human cancers (Petitjean et al., 2007). This means that the gene mutation is observed both in cancer types that are relatively easy to cure as well as in cancer types that are more difficult to cure. In the data set described in Section 4.2 695 patients (44.2%) have mutation in the TP53 gene. Therefore, the positive regression coefficient is very likely caused by the fact that many patients in the data set have the TP53 gene and consequently also a lot of patients with higher survival times have the gene mutation.

There are also variables that are specific for a certain type of cancer. For example, VHL has a relatively large negative coefficient for both breast cancer patients and esophagogastric cancer

patient, while this is not the case for bladder cancer patients. This indicates that genetic testing on VHL gene mutations could be very useful for the prognosis of breast and esophagogastric cancer patients, but not so much for bladder cancer patients. By finding the subgroup-specific influential variables, genetic testing becomes more efficient since it make it possible to focus on a specific gene or set of genes, rather than on all of a person's genes.

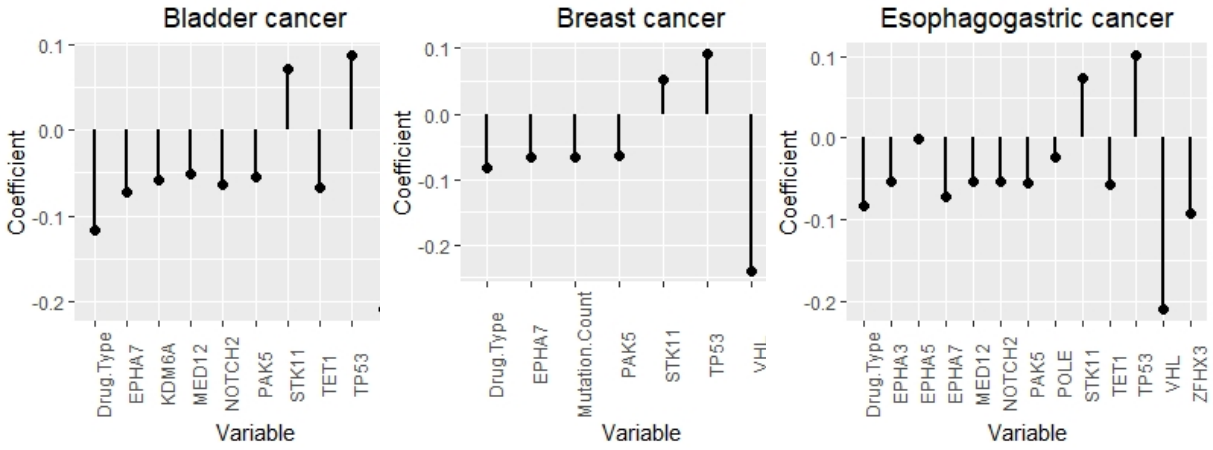

Figure 5.6: Mean estimated regression coefficients (averaged across all training sets) of selected variables. Only the regression coefficients of variables with an absolute value higher than 0.05 are included.

6 Conclusion

The goal of this research was to find an answer to the question: "How can information sharing between subgroups be used to improve the survival prediction of cancer patients based on genetic information?". A novel model is introduced which combines information sharing, lasso regularization and the non-linearity of the Cox's proportional hazard model.

In order to obtain the predictive performance of this new model, the model is compared to eight other models, both linear prediction models and non-linear Cox's proportional hazard models. The models that are compared differ in the way the regression coefficients are estimated. Pooled methods use the observations from all subgroups, but cannot distinguish between different underlying regression models for the different subgroups. Subgroup methods, on the other hand, can estimate different regression coefficients for each subgroup, but perform poorly when the sample size per subgroup is small. The joint methods can use the properties of both the pooled and subgroup method. It estimates different regression coefficients for every subgroup like in the subgroup method, but in addition it can use information from all observations like the pooled method.

All the models were implemented using both simulation data and real data of cancer patients. From the results it follows that information sharing between subgroups is useful, but attention should be paid to the type of dependent variable that is predicted. In this paper's setting, the dependent

variable is survival duration, which means that linear models are inappropriate for prediction due to censoring. Consequently, information sharing in linear models does not improve survival prediction.

On the other hand, when information sharing is applied to non-linear Cox proportional hazard models, as done in the novel joint-LASSO Cox model, the survival prediction gains accuracy according to the results. The model shows the best performance both in the simulation and real data case after comparison of Goodman and Kruskal's γ. The joint-LASSO Cox method, in addition, obtains a relatively high predictive ability for both small subgroup sample sizes as well as for large sample sizes and values in between. This makes the model applicable for varying research settings.

For the real data case, the differences in results between the different models were small. This might be due to the lack of hazard proportionality in the data set. For that reason, it is recommended to verify the proportional hazard assumptions using Schoenfeld residuals (Babińska et al., 2015) before applying the joint-LASSO Cox model. If the data set violates the assumptions, it would be a good idea to further investigate the possibilities of information sharing in other non-linear models. For example, Miladinovic et al. (2012) propose the Royston-Parmar model as an alternative for the Cox model in prognosticating patient survival.

Another aspect that requires further investigation is the amount of fusion between subgroups. In this research, the amount of information sharing between subgroups is based on the assumption that similarity in the features means similarity between underlying regression coefficients. This assumption might not be correct. Therefore, it would be a good idea to look into other ways to define the amount of fusion.

In addition, the only penalization method used in this paper is the LASSO technique since this technique performs both variable selection and regularization. Other papers, like Ogutu et al. (2012), have already comparatively evaluated the predictive performance of multiple regularized linear regression methods. They found that the elastic net, LASSO and adaptive LASSO outperform the adaptive elastic net, ridge regression and ridge regression BLUP. It would be interesting to perform such a comparative study, with different regularization techniques, such as elastic net and adaptive lasso, for the non-linear joint estimation model as well.

References

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statistics in medicine*, *8*(8), 907–925.

Algamal, Z. Y., & Lee, M. H. (2015). Penalized logistic regression with the adaptive lasso for gene selection in high-dimensional cancer classification. *Expert Systems with Applications*, *42*(23), 9326–9332.

Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. (2015). High-dimensional qsar prediction of anticancer potency of imidazo [4, 5-b] pyridine derivatives using adjusted adaptive lasso. *Journal of Chemometrics*, *29*(10), 547–556.

Babińska, M., Chudek, J., Chełmecka, E., Janik, M., Klimek, K., & Owczarek, A. (2015). Limitations of cox proportional hazards analysis in mortality prediction of patients with acute coronary syndrome. *Studies in Logic, Grammar and Rhetoric*, *43*(1), 33–48.

Bair, E., & Tibshirani, R. (2004). Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol*, *2*(4), e108.

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate cox proportional hazards models. *Statistics in medicine*, *24*(11), 1713–1723.

Bhatt, A. N., Mathur, R., Farooque, A., Verma, A., Dwarakanath, B., et al. (2010). Cancer biomarkers-current perspectives. *Indian J Med Res*, *132*(2), 129–149.

Bullinger, L., Döhner, K., Bair, E., Fröhling, S., Schlenk, R. F., Tibshirani, R., ... Pollack, J. R. (2004). Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *New England Journal of Medicine*, *350*(16), 1605–1616.

cBioPortal. (2019). *Genomic and survival data from 1661 tumor-normal pairs from 1661 patients with various cancer types sequenced with the msk-impact assay.* Retrieved from `https://www.cbioportal.org/study/summary?id=tmb_mskcc_2018`

Chen, M., & Kianifard, F. (1999). Application of goodman-kruskal's gamma for ordinal data, in comparing several ordered treatments: A different approach. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *41*(4), 491–498.

Cheng, D. T., Mitchell, T. N., Zehir, A., Shah, R. H., Benayed, R., Syed, A., ... others (2015). Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (msk-impact): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *The Journal of molecular diagnostics*, *17*(3), 251–264.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–202.

Davicioni, E., Anderson, J. R., Buckley, J. D., Meyer, W. H., & Triche, T. J. (2010). Gene expression profiling for survival prediction in pediatric rhabdomyosarcomas: a report from the children's oncology group. *Journal of Clinical Oncology*, *28*(7), 1240.

Decaux, O., Lodé, L., Magrangeas, F., Charbonnel, C., Gouraud, W., Jézéquel, P., ... others (2008). Prediction of survival in multiple myeloma based on gene expression profiles reveals cell cycle and chromosomal instability signatures in high-risk patients and hyperdiploid signatures in low-risk patients: a study of the intergroupe francophone du myelome. *Journal of Clinical Oncology*, *26*(29), 4798–4805.

Dondelinger, F., Mukherjee, S., & Initiative, A. D. N. (2020). The joint lasso: high-dimensional regression for group structured data. *Biostatistics*, *21*(2), 219–235.

Dondelinger, F., & Wilkinson, O. (2018). fuser: Fused lasso for high-dimensional regression over groups. Retrieved from `https://CRAN.R-project.org/package=fuser` (R package version 1.0.1)

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. Retrieved from `http://www.jstatsoft.org/v33/i01/`

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, *49*(268), 732–764.

Hebenstreit, D., & Teichmann, S. A. (2011). Analysis and simulation of gene expression profiles in pure and mixed cell populations. *Physical biology*, *8*(3), 035013.

Huang, E. W., Bhope, A., Lim, J., Sinha, S., & Emad, A. (2020). Tissue-guided lasso for prediction of clinical drug response using preclinical samples. *PLoS computational biology*, *16*(1), e1007607.

Kiefer, N. M. (1988). Economic duration data and hazard functions. *Journal of economic literature*, *26*(2), 646–679.

Le Clerg, E. L. (1957). *Mean separation by the functional analysis of variance and multiple comparisons* (Vol. 3). Agricultural Research Service, US Department of Agriculture.

Lossos, I. S., Czerwinski, D. K., Alizadeh, A. A., Wechser, M. A., Tibshirani, R., Botstein, D., & Levy, R. (2004). Prediction of survival in diffuse large-b-cell lymphoma based on the expression of six genes. *New England Journal of Medicine*, *350*(18), 1828–1837.

Madjar, K. (2020). R source code and data for "weighted cox regression for the prediction of heterogeneous patient subgroups". *GitHub repository*. `https://github.com/KatrinMadjar/WeightedCoxRegression`.

Madjar, K., & Rahnenführer, J. (2020). Weighted cox regression for the prediction of heterogeneous patient subgroups. *arXiv preprint arXiv:2003.08965*.

Miladinovic, B., Kumar, A., Mhaskar, R., Kim, S., Schonwetter, R., & Djulbegovic, B. (2012). A flexible alternative to the cox proportional hazards model for assessing the prognostic accuracy of hospice patient survival. *PloS one*, *7*(10), e47804.

Ng'andu, N. H. (1997). An empirical comparison of statistical tests for assessing the proportional hazards assumption of cox's model. *Statistics in medicine*, *16*(6), 611–626.

Nguyen, D. V., & Rocke, D. M. (2002). Partial least squares proportional hazard regression for application to dna microarray survival data. *Bioinformatics*, *18*(12), 1625–1632.

Ogutu, J. O., Schulz-Streeck, T., & Piepho, H.-P. (2012). Genomic selection using regularized linear regression models: ridge regression, lasso, elastic net and their extensions. *BMC proceedings*, *6*(S2), S10.

Petitjean, A., Mathe, E., Kato, S., Ishioka, C., Tavtigian, S. V., Hainaut, P., & Olivier, M. (2007). Impact of mutant p53 functional properties on tp53 mutation patterns and tumor phenotype: lessons from recent developments in the iarc tp53 database. *Human mutation*, *28*(6), 622–629.

Sermpinis, G., Tsoukas, S., & Zhang, P. (2018). Modelling market implied ratings using lasso variable selection techniques. *Journal of Empirical Finance*, *48*, 19–35.

Shedden, K., Taylor, J. M., Enkemann, S. A., Tsao, M.-S., Yeatman, T. J., Gerald, W. L., ... others (2008). Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nature medicine*, *14*(8), 822.

Simmons, C. P., Koinis, F., Fallon, M. T., Fearon, K. C., Bowden, J., Solheim, T. S., ... Laird, B. J. (2015). Prognosis in advanced lung cancer–a prospective study examining key clinicopathological factors. *Lung Cancer*, *88*(3), 304–309.

Therneau, T. M., & Grambsch, P. M. (2000). The cox model. In *Modeling survival data: extending the cox model* (pp. 39–77). Springer.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tibshirani, R. (1997). The lasso method for variable selection in the cox model. *Statistics in medicine*, *16*(4), 385–395.

Van De Vijver, M. J., He, Y. D., Van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., ... others (2002). A gene-expression signature as a predictor of survival in breast cancer. *New England Journal of Medicine*, *347*(25), 1999–2009.

Xiang, Y., & Chen, H. (2017). The study of credit scoring model based on group lasso. *Procedia computer science*, *122*, 677–684.

7 Appendix

7.1 Simulation

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.314									
Pooled-LASSO	0.442	0.128 (0.026*)								
Subgroup-LASSO	0.433	0.119 (0.030*)	-0.009 (0.866)							
Binary-LASSO	0.490	0.176 (0.003*)	0.048 (0.369)	0.057 (0.317)						
Joint-LASSO	0.273	-0.041 (0.446)	-0.169 (0.005*)	-0.160 (0.006*)	-0.217 (0.000*)					
Joint-LASSO switching parameters	0.130	-0.183 (0.002*)	-0.311 (0.000*)	-0.303 (0.000*)	-0.359 (0.000*)	-0.143 (0.010*)				
Pooled-LASSO Cox	0.526	0.212 (0.001*)	0.084 (0.140)	0.093 (0.115)	0.036 (0.497)	0.253 (0.000*)	0.396 (0.000*)			
Subgroup-LASSO Cox	0.661	0.348 (0.000*)	0.219 (0.000*)	0.228 (0.000*)	0.171 (0.003*)	0.388 (0.000*)	0.531 (0.000*)	0.135 (0.015*)		
Joint-LASSO Cox	0.677	0.363 (0.000*)	0.235 (0.000*)	0.244 (0.000*)	0.187 (0.002*)	0.404 (0.000*)	0.547 (0.000*)	0.151 (0.009*)	0.016 (0.765)	

Table 7.1: Mean γ and differences in mean γ for subgroup 1A.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.292									
Pooled-LASSO	0.538	0.246 (0.003*)								
Subgroup-LASSO	0.353	0.061 (0.401)	-0.185 (0.019*)							
Binary-LASSO	0.516	0.224 (0.005*)	-0.022 (0.762)	0.163 (0.030*)						
Joint-LASSO	0.272	-0.021 (0.776)	-0.267 (0.002*)	-0.082 (0.292)	-0.245 (0.003*)					
Joint-LASSO switching parameters	0.243	-0.049 (0.526)	-0.295 (0.001*)	-0.110 (0.171)	-0.273 (0.001*)	-0.029 (0.695)				
Pooled-LASSO Cox	0.581	0.289 (0.001*)	0.043 (0.553)	0.228 (0.005*)	0.065 (0.402)	0.310 (0.000*)	0.338 (0.000*)			
Subgroup-LASSO Cox	0.651	0.359 (0.000*)	0.113 (0.147)	0.298 (0.000*)	0.135 (0.095)	0.380 (0.000*)	0.408 (0.000*)	0.070 (0.339)		
Joint-LASSO Cox	0.686	0.394 (0.000*)	0.148 (0.066)	0.333 (0.000*)	0.170 (0.040*)	0.415 (0.000*)	0.443 (0.000*)	0.105 (0.177)	0.035 (0.627)	

Table 7.2: Mean γ and differences in mean γ for subgroup 1B.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.206									
Pooled-LASSO	0.330	0.124 (0.055)								
Subgroup-LASSO	0.501	0.295 (0.000*)	0.171 (0.007*)							
Binary-LASSO	0.387	0.181 (0.007*)	0.056 (0.334)	-0.114 (0.055)						
Joint-LASSO	0.241	0.035 (0.548)	-0.089 (0.152)	-0.260 (0.000*)	-0.146 (0.025*)					
Joint-LASSO switching parameters	0.246	0.040 (0.520)	-0.084 (0.152)	-0.255 (0.000*)	-0.141 (0.026*)	0.005 (0.933)				
Pooled-LASSO Cox	0.563	0.357 (0.000*)	0.233 (0.000*)	0.062 (0.288)	0.176 (0.006*)	0.322 (0.000*)	0.317 (0.000*)			
Subgroup-LASSO Cox	0.582	0.375 (0.000*)	0.251 (0.000*)	0.081 (0.196)	0.195 (0.003*)	0.340 (0.000*)	0.335 (0.000*)	0.018 (0.752)		
Joint-LASSO Cox	0.734	0.528 (0.000*)	0.404 (0.000*)	0.233 (0.000*)	0.347 (0.000*)	0.493 (0.000*)	0.488 (0.000*)	0.171 (0.007*)	0.153 (0.012*)	

Table 7.3: Mean γ and differences in mean γ for subgroup 2A.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.341									
Pooled-LASSO	0.365	0.024 (0.701)								
Subgroup-LASSO	0.449	0.108 (0.088)	0.084 (0.136)							
Binary-LASSO	0.348	0.008 (0.899)	-0.016 (0.772)	-0.100 (0.094)						
Joint-LASSO	0.223	-0.118 (0.041*)	-0.141 (0.026*)	-0.226 (0.001*)	-0.125 (0.044*)					
Joint-LASSO switching parameters	0.344	0.003 (0.959)	-0.021 (0.725)	-0.105 (0.090)	-0.005 (0.933)	0.120 (0.046*)				
Pooled-LASSO Cox	0.605	0.264 (0.000*)	0.240 (0.000*)	0.156 (0.008*)	0.256 (0.000*)	0.381 (0.000*)	0.261 (0.000*)			
Subgroup-LASSO Cox	0.638	0.297 (0.000*)	0.273 (0.000*)	0.189 (0.002*)	0.290 (0.000*)	0.415 (0.000*)	0.294 (0.000*)	0.033 (0.552)		
Joint-LASSO Cox	0.772	0.431 (0.000*)	0.407 (0.000*)	0.323 (0.000*)	0.424 (0.000*)	0.549 (0.000*)	0.428 (0.000*)	0.167 (0.006*)	0.134 (0.021*)	

Table 7.4: Mean γ and differences in mean γ for subgroup 2B.
* indicates a significant differences in mean for a 5% significance level

7.2 Real data

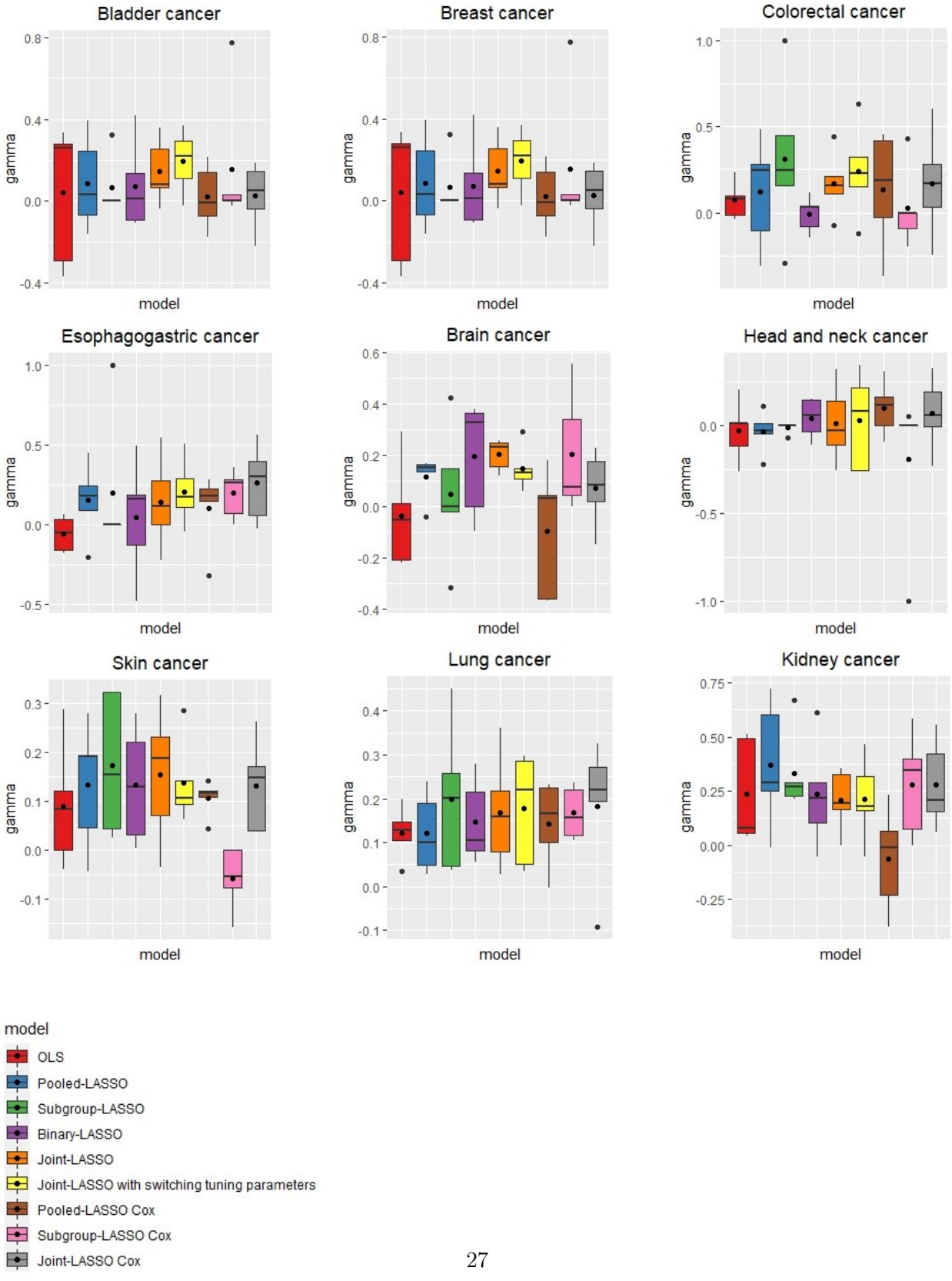

Figure 7.1: Boxplots of γ for the prediction per cancer type.

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.041									
Pooled-LASSO	0.088	0.047 (0.769)								
Subgroup-LASSO	0.066	0.025 (0.863)	-0.022 (0.886)							
Binary-LASSO	0.073	0.032 (0.835)	-0.015 (0.918)	0.007 (0.960)						
Joint-LASSO	0.145	0.104 (0.527)	0.057 (0.696)	0.079 (0.623)	0.071 (0.643)					
Joint-LASSO switching parameters	0.195	0.154 (0.362)	0.107 (0.506)	0.129 (0.440)	0.121 (0.459)	0.050 (0.746)				
Pooled-LASSO Cox	0.020	-0.021 (0.883)	-0.068 (0.677)	-0.046 (0.765)	-0.053 (0.739)	-0.125 (0.454)	-0.175 (0.304)			
Subgroup-LASSO Cox	0.157	0.116 (0.486)	0.069 (0.655)	0.091 (0.579)	0.084 (0.601)	0.012 (0.932)	-0.038 (0.794)	0.137 (0.416)		
Joint-LASSO Cox	0.025	-0.016 (0.910)	-0.063 (0.698)	-0.041 (0.789)	-0.048 (0.762)	-0.120 (0.470)	-0.170 (0.317)	0.005 (0.972)	-0.132 (0.431)	

Table 7.5: Mean γ and differences in mean γ of bladder cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.057									
Pooled-LASSO	0.009	-0.048 (0.855)								
Subgroup-LASSO	-0.170	-0.227 (0.414)	-0.179 (0.494)							
Binary-LASSO	0.160	0.102 (0.706)	0.150 (0.595)	0.330 (0.254)						
Joint-LASSO	0.118	0.060 (0.806)	0.108 (0.690)	0.288 (0.309)	-0.042 (0.872)					
Joint-LASSO switching parameters	0.159	0.101 (0.699)	0.149 (0.591)	0.329 (0.251)	-0.001 (0.996)	0.041 (0.867)				
Pooled-LASSO Cox	0.037	-0.020 (0.934)	0.028 (0.910)	0.207 (0.446)	-0.123 (0.659)	-0.081 (0.758)	-0.122 (0.655)			
Subgroup-LASSO Cox	-0.009	-0.067 (0.807)	-0.019 (0.940)	0.161 (0.513)	-0.169 (0.554)	-0.127 (0.648)	-0.168 (0.553)	-0.046 (0.860)		
Joint-LASSO Cox	0.316	0.258 (0.345)	0.306 (0.277)	0.486 (0.091)	0.156 (0.519)	0.198 (0.459)	0.157 (0.542)	0.278 (0.317)	0.325 (0.253)	

Table 7.6: Mean γ and differences in mean γ of breast cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.074									
Pooled-LASSO	0.120	0.046 (0.791)								
Subgroup-LASSO	0.311	0.238 (0.241)	0.192 (0.338)							
Binary-LASSO	-0.007	-0.080 (0.665)	-0.126 (0.512)	-0.318 (0.124)						
Joint-LASSO	0.169	0.095 (0.620)	0.049 (0.790)	-0.143 (0.458)	0.175 (0.381)					
Joint-LASSO switching parameters	0.242	0.169 (0.399)	0.123 (0.533)	-0.069 (0.690)	0.249 (0.224)	0.073 (0.692)				
Pooled-LASSO Cox	0.132	0.058 (0.754)	0.012 (0.944)	-0.180 (0.361)	0.138 (0.482)	-0.037 (0.831)	-0.110 (0.566)			
Subgroup-LASSO Cox	0.028	-0.046 (0.793)	-0.092 (0.621)	-0.284 (0.167)	0.035 (0.842)	-0.141 (0.474)	-0.214 (0.290)	-0.104 (0.589)		
Joint-LASSO Cox	0.168	0.094 (0.638)	0.048 (0.803)	-0.144 (0.474)	0.175 (0.403)	-0.001 (0.996)	-0.074 (0.701)	0.036 (0.842)	0.140 (0.496)	

Table 7.7: Mean γ and differences in mean γ of colorectal cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	-0.060									
Pooled-LASSO	0.152	0.212 (0.271)								
Subgroup-LASSO	0.200	0.260 (0.189)	0.048 (0.789)							
Binary-LASSO	0.045	0.104 (0.538)	-0.107 (0.567)	-0.155 (0.425)						
Joint-LASSO	0.141	0.201 (0.285)	-0.011 (0.948)	-0.059 (0.752)	0.096 (0.594)					
Joint-LASSO switching parameters	0.204	0.264 (0.187)	0.052 (0.781)	0.004 (0.982)	0.159 (0.419)	0.063 (0.742)				
Pooled-LASSO Cox	0.101	0.161 (0.374)	-0.051 (0.778)	-0.099 (0.605)	0.056 (0.739)	-0.04 (0.814)	-0.103 (0.597)			
Subgroup-LASSO Cox	0.195	0.255 (0.193)	0.043 (0.800)	-0.005 (0.975)	0.150 (0.434)	0.054 (0.765)	-0.009 (0.959)	0.094 (0.617)		
Joint-LASSO Cox	0.259	0.319 (0.124)	0.107 (0.588)	0.059 (0.752)	0.214 (0.296)	0.118 (0.556)	0.055 (0.753)	0.158 (0.437)	0.064 (0.740)	

Table 7.8: Mean γ and differences in mean γ of esophagogastric cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	-0.035									
Pooled-LASSO	0.115	0.151 (0.256)								
Subgroup-LASSO	0.047	0.082 (0.508)	-0.069 (0.580)							
Binary-LASSO	0.195	0.231 (0.107)	0.080 (0.560)	0.149 (0.290)						
Joint-LASSO	0.203	0.238 (0.104)	0.087 (0.540)	0.156 (0.281)	0.007 (0.956)					
Joint-LASSO switching parameters	0.148	0.184 (0.191)	0.033 (0.803)	0.102 (0.459)	-0.047 (0.704)	-0.054 (0.692)				
Pooled-LASSO Cox	-0.096	-0.060 (0.626)	-0.211 (0.125)	-0.143 (0.282)	-0.291 (0.045)	-0.298 (0.044)	-0.244 (0.088)			
Subgroup-LASSO Cox	0.203	0.238 (0.101)	0.087 (0.535)	0.156 (0.276)	0.007 (0.954)	0.000 (0.998)	0.054 (0.682)	0.298 (0.042)		
Joint-LASSO Cox	0.071	0.106 (0.413)	-0.044 (0.716)	0.024 (0.842)	-0.124 (0.356)	-0.132 (0.348)	-0.077 (0.552)	0.167 (0.217)	-0.131 (0.341)	

Table 7.9: Mean γ and differences in mean γ of brain cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	-0.032									
Pooled-LASSO	-0.035	-0.004 (0.979)								
Subgroup-LASSO	-0.015	0.017 (0.905)	0.021 (0.891)							
Binary-LASSO	0.041	0.073 (0.653)	0.077 (0.642)	0.056 (0.725)						
Joint-LASSO	0.012	0.043 (0.776)	0.047 (0.766)	0.026 (0.854)	-0.030 (0.847)					
Joint-LASSO switching parameters	0.026	0.058 (0.716)	0.062 (0.705)	0.041 (0.790)	-0.015 (0.916)	0.014 (0.920)				
Pooled-LASSO Cox	0.099	0.131 (0.434)	0.135 (0.425)	0.114 (0.492)	0.058 (0.706)	0.087 (0.591)	0.073 (0.646)			
Subgroup-LASSO Cox	-0.189	-0.158 (0.304)	-0.154 (0.287)	-0.175 (0.272)	-0.231 (0.169)	-0.201 (0.217)	-0.216 (0.193)	-0.289 (0.092)		
Joint-LASSO Cox	0.067	0.099 (0.548)	0.103 (0.538)	0.082 (0.614)	0.026 (0.856)	0.056 (0.726)	0.041 (0.787)	-0.032 (0.825)	0.257 (0.130)	

Table 7.10: Mean γ and differences in mean γ of head and neck cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.091									
Pooled-LASSO	0.134	0.044 (0.584)								
Subgroup-LASSO	0.174	0.083 (0.313)	0.040 (0.609)							
Binary-LASSO	0.133	0.042 (0.587)	-0.001 (0.985)	-0.041 (0.606)						
Joint-LASSO	0.155	0.064 (0.435)	0.020 (0.786)	-0.019 (0.781)	0.022 (0.781)					
Joint-LASSO switching parameters	0.138	0.047 (0.556)	0.004 (0.955)	-0.036 (0.633)	0.005 (0.944)	-0.016 (0.816)				
Pooled-LASSO Cox	0.107	0.016 (0.822)	-0.028 (0.720)	-0.067 (0.409)	-0.026 (0.723)	-0.048 (0.552)	-0.032 (0.689)			
Subgroup-LASSO Cox	-0.057	-0.148 (0.040*)	-0.192 (0.018*)	-0.231 (0.006*)	-0.190 (0.017*)	-0.212 (0.011*)	-0.196 (0.018*)	-0.164 (0.031*)		
Joint-LASSO Cox	0.132	0.041 (0.580)	-0.002 (0.978)	-0.042 (0.604)	-0.001 (0.991)	-0.022 (0.779)	-0.006 (0.938)	0.026 (0.714)	0.190 (0.016*)	

Table 7.11: Mean γ and differences in mean γ of skin cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.122									
Pooled-LASSO	0.121	-0.001 (0.986)								
Subgroup-LASSO	0.199	0.076 (0.382)	0.078 (0.377)							
Binary-LASSO	0.147	0.025 (0.757)	0.026 (0.754)	-0.052 (0.545)						
Joint-LASSO	0.168	0.046 (0.584)	0.047 (0.580)	-0.030 (0.713)	0.022 (0.785)					
Joint-LASSO switching parameters	0.178	0.056 (0.516)	0.057 (0.511)	-0.021 (0.793)	0.031 (0.705)	0.009 (0.898)				
Pooled-LASSO Cox	0.143	0.021 (0.777)	0.022 (0.778)	-0.055 (0.522)	-0.004 (0.962)	-0.025 (0.760)	-0.035 (0.681)			
Subgroup-LASSO Cox	0.167	0.045 (0.586)	0.046 (0.584)	-0.032 (0.708)	0.020 (0.785)	-0.001 (0.986)	-0.011 (0.892)	0.024 (0.764)		
Joint-LASSO Cox	0.183	0.061 (0.481)	0.062 (0.475)	-0.015 (0.837)	0.037 (0.664)	0.015 (0.851)	0.005 (0.942)	0.040 (0.640)	0.016 (0.844)	

Table 7.12: Mean γ and differences in mean γ of lung cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Mean	**Differences in mean**								
		OLS	Pooled-LASSO	Subgroup-LASSO	Binary-LASSO	Joint-LASSO	Joint-LASSO switching parameters	Pooled-LASSO Cox	Subgroup-LASSO Cox	Joint-LASSO Cox
OLS	0.236									
Pooled-LASSO	0.371	0.135 (0.410)								
Subgroup-LASSO	0.333	0.096 (0.546)	-0.038 (0.791)							
Binary-LASSO	0.234	-0.002 (0.988)	-0.137 (0.410)	-0.099 (0.546)						
Joint-LASSO	0.208	-0.028 (0.860)	-0.163 (0.337)	-0.125 (0.458)	-0.026 (0.866)					
Joint-LASSO switching parameters	0.213	-0.023 (0.882)	-0.157 (0.349)	-0.119 (0.473)	-0.021 (0.887)	0.005 (0.970)				
Pooled-LASSO Cox	-0.066	-0.302 (0.065)	-0.437 (0.011*)	-0.399 (0.020*)	-0.300 (0.061)	-0.274 (0.063)	-0.280 (0.072)			
Subgroup-LASSO Cox	0.280	0.044 (0.774)	-0.090 (0.557)	-0.052 (0.718)	0.046 (0.771)	0.072 (0.663)	0.067 (0.682)	0.347 (0.040*)		
Joint-LASSO Cox	0.280	0.044 (0.761)	-0.091 (0.570)	-0.052 (0.733)	0.046 (0.764)	0.072 (0.659)	0.067 (0.676)	0.346 (0.038*)	0.000 (0.998)	

Table 7.13: Mean γ and differences in mean γ of kidney cancer patients.
* indicates a significant differences in mean for a 5% significance level

	Bladder	**Breast**	**Colorectal**	**Esophagogastric**	**Brain**	**Head and neck**	**Skin**	**Lung**	**Kidney**
Drug.Type	-0.116	-0.082	-0.135	-0.084	-0.149	-0.105	-0.147	-0.094	-0.123
Mutation.Count	0.000	-0.066	0.000	0.000	-0.088	-0.004	-0.035	0.000	-0.036
Sex	0.000	-0.006	0.000	0.000	-0.015	-0.003	0.000	0.000	-0.027
TMB.Score	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tumor.Purity	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TP53	0.087	0.091	0.038	0.100	0.055	0.100	0.037	0.101	0.091
TERT	-0.012	0.000	-0.022	-0.033	0.000	-0.007	0.000	-0.029	0.000
KMT2D	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
KRAS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PIK3CA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ARID1A	0.000	-0.008	0.000	-0.001	0.000	-0.012	0.000	0.000	-0.010
NF1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PTPRT	0.000	0.000	-0.003	0.000	0.000	0.000	-0.009	0.000	0.000
BRAF	-0.020	-0.010	-0.045	0.000	-0.051	-0.023	-0.064	0.000	-0.035
FAT1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
KMT2C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PREX2	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000
APC	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PTPRD	-0.048	-0.028	-0.069	-0.042	-0.023	-0.038	-0.036	-0.042	-0.031
PBRM1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PTEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CDKN2A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SET2D	-0.037	-0.035	0.000	-0.042	-0.006	-0.042	-0.008	-0.041	-0.029
GRIN2A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SMARCA4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ATRX	-0.004	0.000	-0.023	0.000	0.000	0.000	0.000	0.000	0.000
ZFHX3	-0.088	-0.063	-0.032	-0.093	-0.046	-0.081	-0.050	-0.096	-0.069
VHL	-0.209	-0.240	0.000	-0.209	-0.240	-0.236	-0.147	-0.209	-0.262
ROS1	-0.012	0.000	-0.010	0.000	-0.025	0.000	-0.061	0.000	-0.006
KMT2A	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOTCH1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EGFR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ERBB4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOTCH4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ARID2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CREBBP	-0.036	0.000	-0.044	-0.029	0.000	-0.017	0.000	-0.028	0.000
PAK5	-0.054	-0.064	-0.022	-0.055	-0.046	-0.067	-0.035	-0.057	-0.067
ATM	-0.036	-0.012	-0.007	-0.028	0.000	-0.027	0.000	-0.031	-0.014
RB1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NRAS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BRCA2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ALK	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EPHA3	-0.037	-0.041	0.000	-0.054	-0.001	-0.047	0.000	-0.053	-0.038
FLT1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
STK11	0.072	0.051	0.048	0.073	0.001	0.067	0.008	0.074	0.048
NOTCH3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SPEN	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
KDR	0.000	0.000	0.000	0.000	0.000	0.000	-0.002	0.000	0.000
KEAP1	0.046	0.033	0.047	0.043	0.004	0.037	0.000	0.045	0.028
PIK3CG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TP63	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MGA	0.000	0.000	0.000	-0.001	0.000	0.000	0.000	0.000	0.000
EPHA7	-0.072	-0.067	-0.005	-0.073	-0.052	-0.080	-0.033	-0.072	-0.076
POLE	-0.014	-0.024	0.000	-0.023	-0.017	-0.024	0.000	-0.022	-0.029

	Bladder	**Breast**	**Colorectal**	**Esophagogastric**	**Brain**	**Head and neck**	**Skin**	**Lung**	**Kidney**
EPHA5	0.000	0.000	0.000	-0.002	0.000	0.000	0.000	-0.003	0.000
EP300	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NOTCH2	-0.064	-0.038	-0.038	-0.053	-0.028	-0.052	-0.032	-0.054	-0.045
PIK3C2G	-0.030	-0.021	0.000	-0.027	-0.015	-0.031	-0.006	-0.028	-0.028
CARD11	0.000	0.000	0.000	0.000	-0.019	0.000	-0.020	0.000	-0.006
RBM10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PTPRS	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TET2	0.000	0.000	0.000	0.000	-0.011	-0.002	0.000	0.000	-0.019
ARID1B	-0.003	-0.020	0.000	-0.007	-0.010	-0.023	0.000	-0.008	-0.037
PDGFRA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
KDM6A	-0.058	-0.026	-0.044	-0.041	-0.003	-0.041	0.000	-0.044	-0.031
DOT1L	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ATR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NSD1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FLT4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
KDM5C	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FBXW7	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ERBB3	-0.004	0.000	0.000	-0.004	0.000	-0.006	0.000	-0.003	-0.003
NTRK3	-0.030	-0.014	-0.019	-0.033	0.000	-0.027	0.000	-0.034	-0.012
CTNNB1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SF3B1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MTOR	-0.003	-0.001	0.000	0.000	-0.016	-0.010	-0.004	0.000	-0.021
ERBB2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
NCOR1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PGR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-0.002	0.000
CIC	0.000	0.000	-0.050	0.000	0.000	0.000	0.000	0.000	0.000
BCOR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MDC1	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IGF1R	0.000	0.000	0.000	0.000	0.000	0.000	-0.018	0.000	0.000
TSC2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
EPHB1	-0.027	-0.003	-0.035	-0.016	0.000	-0.018	0.000	-0.018	-0.003
FGFR3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TSC1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BAP1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ANKRD11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
HGF	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MED12	-0.052	-0.022	-0.011	-0.054	0.000	-0.043	0.000	-0.056	-0.021
ASXL1	-0.026	-0.036	0.000	-0.038	-0.003	-0.042	0.000	-0.036	-0.033
SMAD4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MET	-0.044	-0.026	-0.003	-0.048	0.000	-0.041	0.000	-0.049	-0.024
DICER1	-0.002	0.000	-0.003	0.000	-0.011	-0.010	-0.022	0.000	-0.018
RNF43	-0.019	0.000	-0.112	0.000	0.000	0.000	0.000	0.000	0.000
BRCA1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
GLI1	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.000	0.000
FLT3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
IL7R	-0.016	0.000	0.000	-0.013	0.000	-0.009	0.000	-0.013	-0.004
NTRK1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
STAG2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TET1	-0.068	-0.040	-0.105	-0.057	-0.032	-0.060	-0.070	-0.059	-0.046
PDGFRB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TBX3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 7.14: Mean estimated regression coefficients (averaged across all training sets) of selected variables for every cancer type.