

A Regularized Cox Regression Approach to the Health Evaluation and Linkage to Primary Care (HELP) Clinical Trial

June 25, 2021

Abstract: Substance abuse is a pressing public health problem. Practical linkage of primary medical care to patients undergoing substance abuse treatment could improve patient outcomes, as primary care physicians can play an important role in helping individuals seek out long term treatment. The Health Evaluations and Linkage to Primary Care (HELP) study was a clinical trial designed to evaluate the effect of an experimental intervention and other covariates on whether substance abuse patients seek out primary care. An elastic-net-regularized cox regression model was fit to HELP data to identify features relevant to the primary endpoint: time to linkage with a primary care physician. The elastic net penalty generated a highly parsimonious model that achieved good performance on held out data. The features selected by the elastic net model will hopefully aid in the design of future interventions that encourage primary care linkage in substance abuse patients.

Introduction

Finding ways to practically link medical care and substance abuse treatment is a goal of public health, as many patients with addictions do not receive primary medical care. For example, of those undergoing substance abuse treatment in Boston, only 41% had a primary care physician (Saitz, Mulvey, and Samet 1997). Indeed, patients with substance abuse problems are common in general medicine practice across all demographics, and primary care physicians can play a powerful role in helping patients accept treatment (Weaver *et al.* 1999). For example, substance abuse patients who received regular primary care are less likely to be hospitalized (Laine *et al.* 2001), posing benefits for patients and reducing economic burden on the health care system. Therefore, finding ways to involve primary care physicians in the rehabilitation and chemical dependency treatment process could potentially lead to better outcomes for substance abuse patients and improvements in public health.

The Health Evaluation and Linkage to Primary Care (HELP) study (Samet *et al.* 2003) was a clinical trial in adult patients without a primary care physician who were undergoing in-patient detoxification treatment for alcohol, heroine, and cocaine addiction in the Boston area. Patients were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the primary endpoint being whether the patient attended an appointment with a primary care physician within 12 months. Patients enrolled in the study were Spanish/English speaking adults that had reported alcohol, heroin, or cocaine as their primary or secondary drug of choice. Patients needed to reside in the proximity of the primary care clinic to which they would be referred or were homeless. Patients were excluded that had established primary care relationships, significant dementia, plans to leave the Boston area that would lead to loss of follow up, failed to provide contact information, or were pregnant. Subjects were interviewed at baseline during their detoxification stay and follow-up interviews were undertaken every 6 months for 2 years. A variety of covariates and outcomes were measured per individual.

Methods

The Elastic Net and Cox Survival Model

The elastic net (Zou & Hastie 2004) is a regularization and variable selection method for regression. The elastic net can improve generalization performance (i.e., accurate predictions on future data) and model interpretability (i.e., more parsimonious models) in regression scenarios by inducing model parsimony via variable selection. The Elastic net is a weighted average of the ridge (L_2) and LASSO (L_1) penalties (Supp. Figure 1). Elastic net improves upon the ridge penalty by encouraging model parsimony, i.e., by shrinking unimportant variables to have 0-valued regression coefficients. Elastic net improves upon LASSO by permitting selection of more than n predictors (where n is the number of observations) and by its ability to select groups of correlated predictors (while LASSO randomly selects one predictor from a correlated group). Elastic net is therefore preferable to LASSO in high dimensional settings, where the number of predictors, p , outnumbers the number of observations, n , and there exists correlation structures among the covariates.

The Cox Proportional Hazards (Cox PH) model assumes that the hazard ratio (instantaneous relative risks of experiencing an event) associated with each covariate is fixed; i.e., the covariate's effect does not depend on time. Cox PH assumes a semi-parametric form for the hazard: $h_i(t) = h_0(t)e^{x_i^T \beta}$, where $h_i(t)$ denotes the hazard for patient i at time t , $h_0(t)$ is the baseline hazard at time t , and β is a vector of predictors (the log hazard ratios, length p).

Typically, β is estimated in the Cox PH model by maximizing the *partial likelihood*:

$$L(\beta) = \prod_{i=1}^m \frac{e^{x_{j(i)}^T \beta}}{\sum_{j \in R_i} e^{x_{j(i)}^T \beta}}$$

where $(i \in \{1, \dots, m\})$ denote the observed event times, R_i is the set of indices, j , with $y_j \geq t_i$. The partial likelihood is the product over the event times (i) of conditional probabilities of witnessing the observed failure given one failure occurred among all susceptible individuals (R_i) at time t_i .

Simon *et al.* (2011) noted that maximizing the partial likelihood is equivalent to maximizing a scaled version of the log partial likelihood (since log is a monotonic transformation),

$$\frac{2}{n} \ell(\beta) = \frac{2}{n} \left[\sum_{i=1}^m x_{j^{(i)}}^T \beta - \log \left(\sum_{j \in R_i} e^{x_j^T \beta} \right) \right]$$

Next, consider the **elastic net penalty** on the vector of β :

$$\lambda P_\alpha(\beta) = \lambda \left(\alpha \sum_{k=1}^p |\beta_k| + \frac{1}{2}(1 - \alpha) \sum_{k=1}^p \beta_k^2 \right)$$

where $\alpha \in [0, 1]$ denotes the relative weights of the L_1 and L_2 penalties¹ and λ denotes the regularization strength. Incorporating the elastic net penalty into the partial likelihood yields the objective function from which penalized $\hat{\beta}$ values can be obtained:

$$\hat{\beta} = \operatorname{argmax}_\beta \left[\frac{2}{n} \left(\sum_{i=1}^m x_{j^{(i)}}^T \beta - \log \left(\sum_{j \in R_i} e^{x_j^T \beta} \right) \right) - \lambda P_\alpha(\beta) \right]$$

I implemented the regularized Cox model fitting using the `glmnet` package (Friedman *et al.* 2010) (Simon *et al.* 2011) as described in the Coxnet vignette (Tay *et al.* 2021).

Data Cleaning

I split the dataset into training and test sets. I filtered the training dataset consisting of 347 patients/observations (rows) and 788 features (columns) for features with fewer than 10% missing values. The filtered dataset contained 347 observations and 503 features. Categorical features (with minimum value of 0 or 1, a maximum value less than 16, and no decimal values) were encoded as factors. Continuous features were treated as numeric variables, and NA values were imputed by taking the mean of all the values in the column and rounding to the nearest integer. I also excluded the `e14e` feature (Have you had biofeedback in the last 6 months?) from the model, because every respondent replied “No”, and a cox model cannot be fit to a factor with only 1 level.

Model tuning

The α parameter specifying the relative weights of the L_1 and L_2 penalties was tuned. In order to evaluate the performance of the coxnet models while varying α and λ (the regularization strength), I performed 10-fold cross validation (CV) over a logarithmically spaced grid of λ values while varying the α parameter over $\{0, 0.25, 0.5, 0.75, 1\}$ using the `cv.glmnet` function in the package `glmnet`. Parameters were tuned to minimize the CV deviance.

Shown in Figures 2 and 3, choice of α largely had no effect on the deviance and C statistics produced in the cross validation experiment: for all $\alpha > 0$, the curves achieved similar deviances and C-statistics in the CV experiment. The one exception was $\alpha = 0$, where the C statistic was much lower cross most $\log(\lambda)$ values surveyed. Thus, any $\alpha > 0$ yielded comparable performance in CV experiments. I elected to use an equal weighting of the LASSO and Ridge penalties ($\alpha = 0.5$) for my model.

¹ $\alpha = 1$ returns the LASSO penalty, $\alpha = 0$ returns ridge penalty.

To tune the regularization strength λ , I performed a 10-fold CV experiment (with $\alpha = 0.5$) and recorded the optimal λ values that produced the smallest deviance. I selected the maximum optimal regularization strength, $\lambda = 0.171$, to prioritize parsimony in my final model.

Results

The elastic net penalty produced a parsimonious cox model with 8 predictors, representing an approximately 63-fold reduction in model complexity relative to the full model. The predictors include: **a11a** (a binary variable denoting whether a patient currently had a living mother), **b3g** (a categorical variable denoting whether a patient believed their health limited them in walking > 1 mile), **b3h** (a categorical variable denoting whether a patient believed their health limited them in walking several blocks), **f1j** (a categorical variable denoting whether a patient is fearful), **o1d** (a categorical variable denoting the number of people who supported the patient’s abstinence from drugs/alcohol), **group** (a binary variable specifying assignment to either the treatment group of control arm of the trial), **h3_prb** (a binary variable specifying whether an individual has a substance problem related to heroin), and **inter** (categorical variable denoting interpersonal consequences of drug use). The regularized estimates and hazard ratios are shown in Table 1. The cox model achieved a C-statistic of 0.667 when applied to 100 observations in the test dataset, indicating that elastic-net-regularized Cox model was better than random chance at distinguishing between events (those who linked with a primary care physician in 12 months) and non-events (those who did not link in 12 months) on the basis of risk.

Discussion

The Health Evaluation and Linkage to Primary Care (HELP) study was a clinical trial that sought to test an intervention and identify factors related to linking drug detoxification programs to primary medical care. Patients were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the primary endpoint being whether the patient attended an appointment with a primary care physician within 12 months. The right censored time-to-event data format made survival analysis methods preferable for the analysis at hand.

The HELP study represented high-dimensional data, as the number of predictors (788) in the HELP study exceeded the number of observations (347). To ensure model interpretability and good generalization performance, I imposed an elastic net penalty on the model’s partial likelihood. I employed a cross validation scheme to tune the α and λ hyperparameters, specifying the weight of the L_1 and L_2 penalties and regularization strength respectively. When I fit a Cox PH model with the tuned hyperparameters, I obtained a parsimonious model with 8 covariates, representing a nearly 63-fold reduction in model complexity relative to the full model. The model included features involving overall health (**b3g** and **b3h**), psychological state (**f1j**), and interpersonal relationships (**a11a**, **h3_prb**, **o1d**, and **inter**). The covariates included the randomization group, illustrating that the experimental treatment was highly determinative of primary care linkage. These results could pose real benefits to the outcomes of substance abuse patients.

One key limitation of this study was the **glmnet** package does not support diagnostic plots to test the proportional hazards assumption. To indirectly address this question, I fit a unregularized Cox model and used the **cox.zph** function to test for significant deviations from the proportional hazards assumptions. Only **group** and **b3g** showed a significant departure from PH at the $p < 0.05$ level (Table 2). Yet a diagnostic plot of $\log(-\log(S(t)))$ over time² illustrates that the groups display approximately parallel curves, indicative that any violation of the proportional hazards assumption is relatively minor (Figure 3).

In summary, using a cox model with an elastic net penalty, I identified a parsimonious model for drug rehab patients connecting with primary physicians in the Health Evaluation and Linkage to Primary Care study. The final model, **final_fit**, contains 8 features, illuminating what covariates are predictive of primary care

²Where $S(t) = P(T \geq t)$ = prob. of surviving until after time t is the survival function.

linkage in drug rehabilitation patients. These covariates could be potentially relevant to designing future interventions to encourage primary care linkage in substance abuse patients.

References

1. Laine, C., Hauck, W. W., Gourevitch, M. N., Rothman, J., Cohen, A., & Turner, B. J. Regular outpatient medical and drug abuse care and subsequent hospitalization of persons who use illicit drugs. *JAMA* 285, 2355-2362, (2001).
2. Samet, J. H., Larson, M. J., Horton, N. J., Doyle, K., Winter, M., Saitz, R. Linking alcohol-and drug-dependent adults to primary medical care: a randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction*, 98, 509-16 (2003).
3. Saitz, R., Mulvey, K. P. & Samet, J. H. The substance-abusing patient and primary care: Linkage via the addiction treatment system? *Subst. Abus.* 18, 187–195 (1997).
4. Simon, N., Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent. *J. Stat. Softw.* 39, (2011).
5. Tay, K., Simon, N., Friedman, J., Hastie, T., Tibshirani, R., & Narasimhan, B. Regularized Cox Regression. (2021).
6. Weaver, M. F., Jarvis, M. A. E. & Schnoll, S. H. Role of the Primary Care Physician in Problems of Substance Abuse. *JAMA Intern. Med.* 159, 913–924 (1999).
7. Zou, H. & Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc.* 67, 301–320 (2004).
8. Friedman, J., Hastie, T., Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* 33, 1-22 (2010).

Supplemental Figures

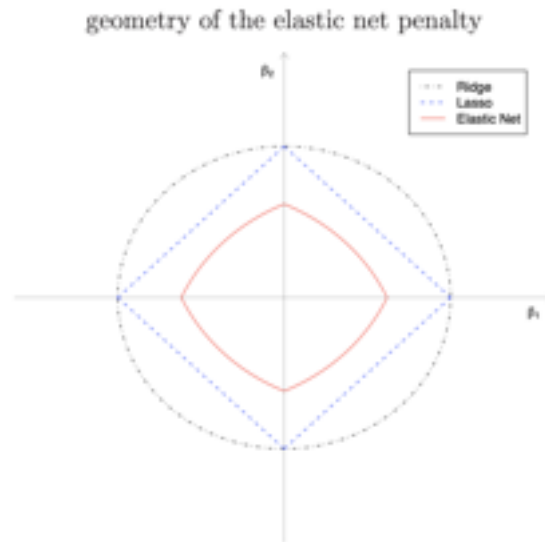


Figure 1: Figure borrowed from Zou and Hastie (2004). Figure illustrates the constraint space of LASSO (blue diamond), Ridge (black circle), and Elastic Net (red deformed diamond). The geometry of the elastic net constraint region encourages both variable shrinkage and model sparsity, combining the benefits of LASSO and Ridge.

Table 1: Variables retained in the Cox model after application of the elastic net penalty. Each term is paired with its shrunken regression coefficient and its shrunken Hazard Ratio (HR).

term	estimate	HR
a11a1	-0.0692133	0.9331276
b3g3	-0.0309744	0.9695004
b3h3	-0.0448149	0.9561745
f1j2	0.0077324	1.0077624
o1d5	0.0069187	1.0069427
group1	0.7954306	2.2153947
h3_prb1	-0.0928621	0.9113192
inter3	0.0858470	1.0896396

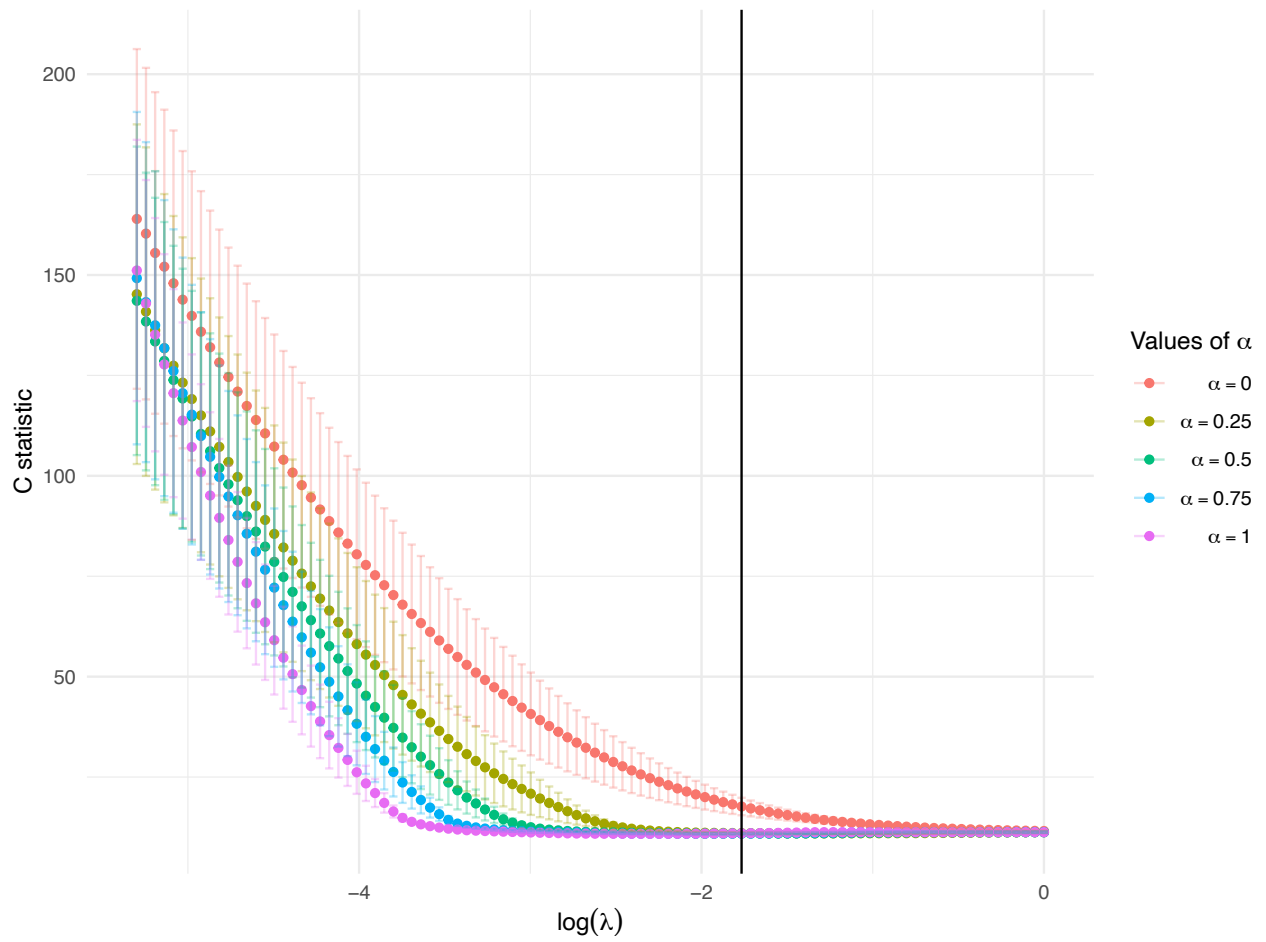


Figure 2: 10-fold CV deviances associated with different regularization strengths (λ) and different elastic net penalty weights (α). The vertical black line denotes the optimal choice of λ , the regularization strength.

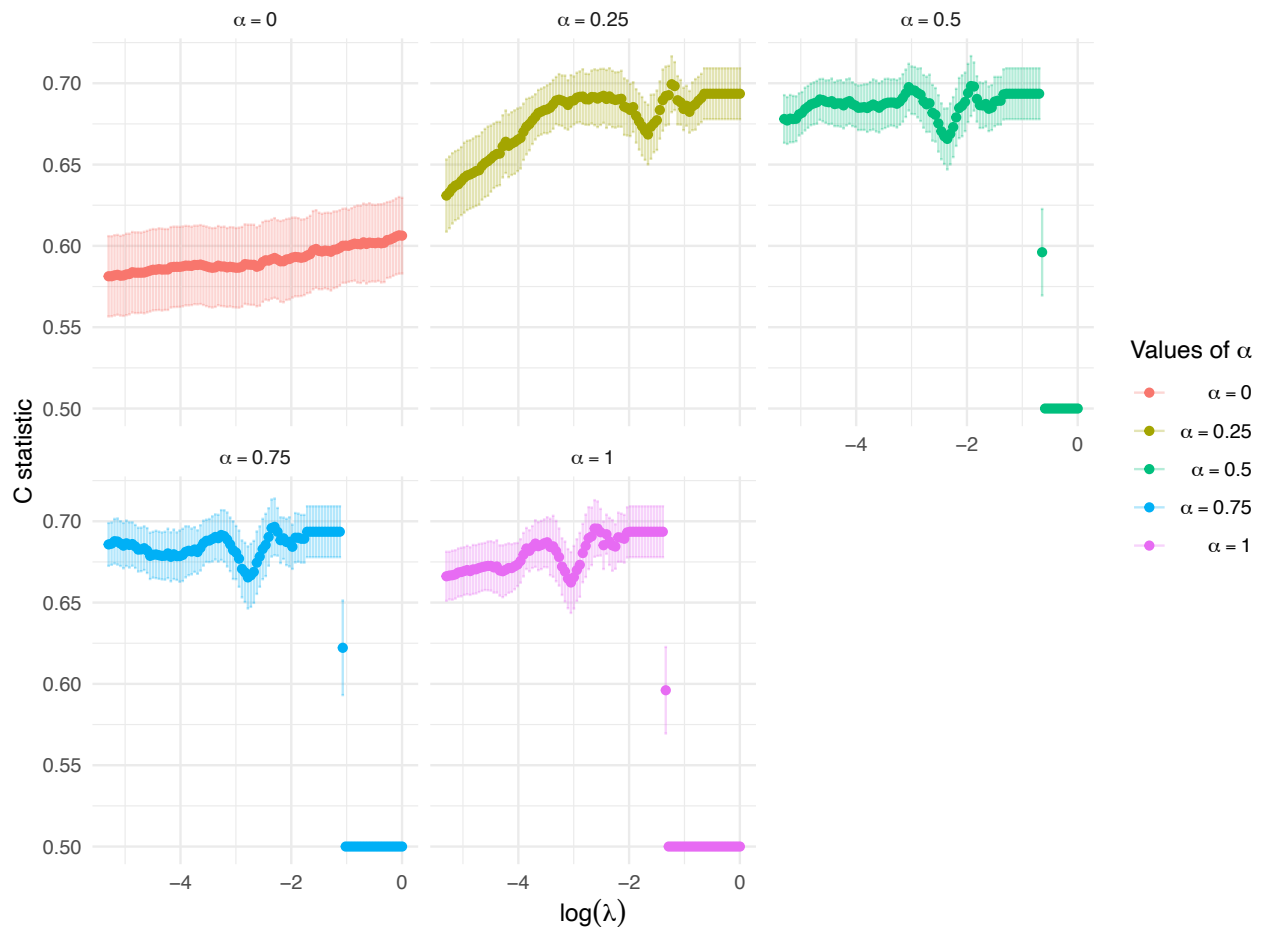


Figure 3: 10-fold CV C-statistics associated with different regularization strengths (λ) and different elastic net penalty weights (α). Lasso ($\alpha=0$) demonstrated the lowest C-statistic for all models assayed, while any $\alpha>0$ produced similar profiles of CV C-statistic.

Table 2: Results of testing proportional hazards assumption for 8 covariates identified by the Elastic Net Cox Model.

	chisq	df	p
a11a	1.5400668	1	0.2146079
b3g	4.3031187	1	0.0380425
b3h	0.4600214	1	0.4976140
flj	2.0248626	1	0.1547429
o1d	1.9291961	1	0.1648464
group	7.1760422	1	0.0073884
h3_prb	0.4330590	1	0.5104917
inter	1.0497541	1	0.3055637
GLOBAL	17.6587245	8	0.0239358

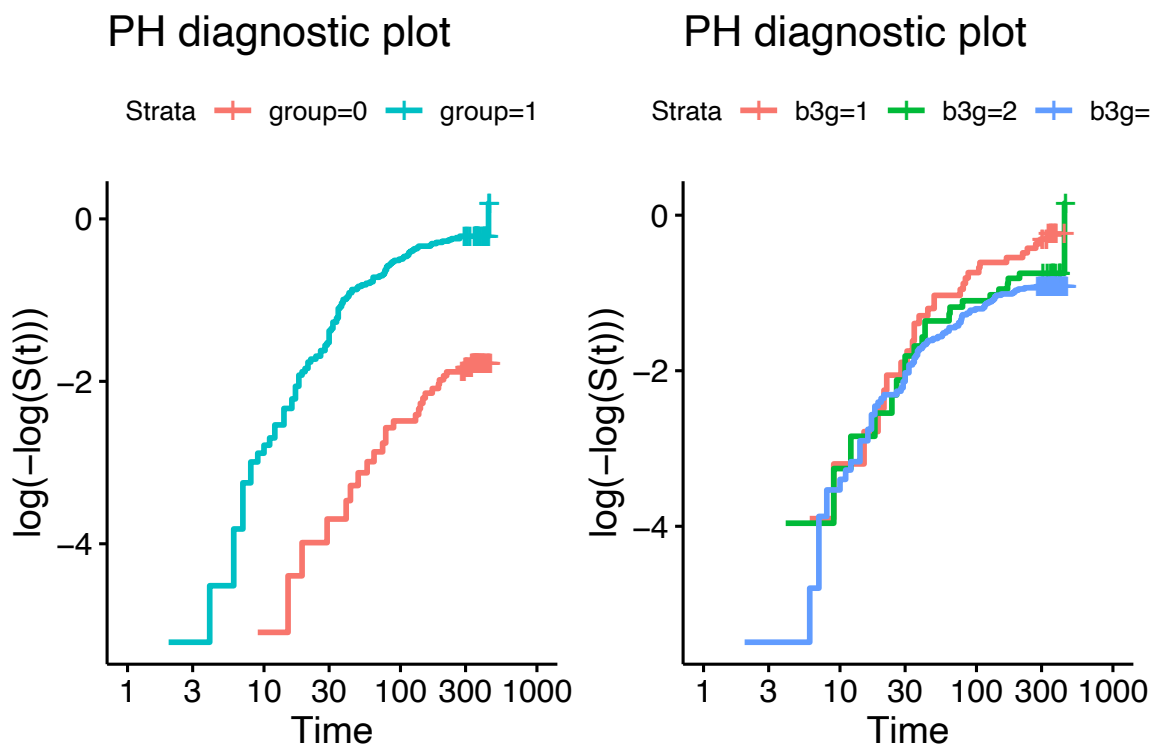


Figure 4: log-log diagnostic plot of unregularized Cox model fit to the ‘group’ and ‘b3g’ variables demonstrates that proportional hazards (PH) assumption is not greatly violated.