Gains of Complicated Distress Model: Danish Evidence

Joint work with Pia Mølgaard.

CBS COPENHAGEN BUSINESS SCHOOL
HANDELSHØJSKOLEN

Motivation

Firm level models are important when evaluating riskiness of portfolios.

Models need to **rank** firms in terms of riskiness.

Models need to be **well calibrated**.
Particularly, in the cross section of firms.

Complex models may improve on either.
Gains may be large.

Talk overview

Problem and data.

Four models.

1. Generalized linear model (logit).
2. Generalized additive model.
3. Gradient tree boosting.
4. Generalized linear mixed model.

Main conclusions

There are gains of more complex models on some metrics.

The gains are smaller than suggested in recent papers.

The more complex models does not overcome issues with temporal effects.
Temporal effects: changes over time.

The temporal effects can be addressed with random effects models.

Problem and data

All ApS and A/S.

Financial statements from Experian and Bisnode.

Event: enters distress period.

In progress of or exited due to bankruptcy or compulsory dissolution ("tvangsopløsning").

Winsorize ratios with denominator

$$\text{size}_{it} = \max(\text{equity}_{it}, 0) + \text{debt}_{it}$$

Similar issue as in Campbell and Szilagyi (2008).

$44\,242$ distresses, $194\,684$ firms and $1\,322\,068$ firm years.

What is an event?

Discrete hazard models

as in Shumway (2001).

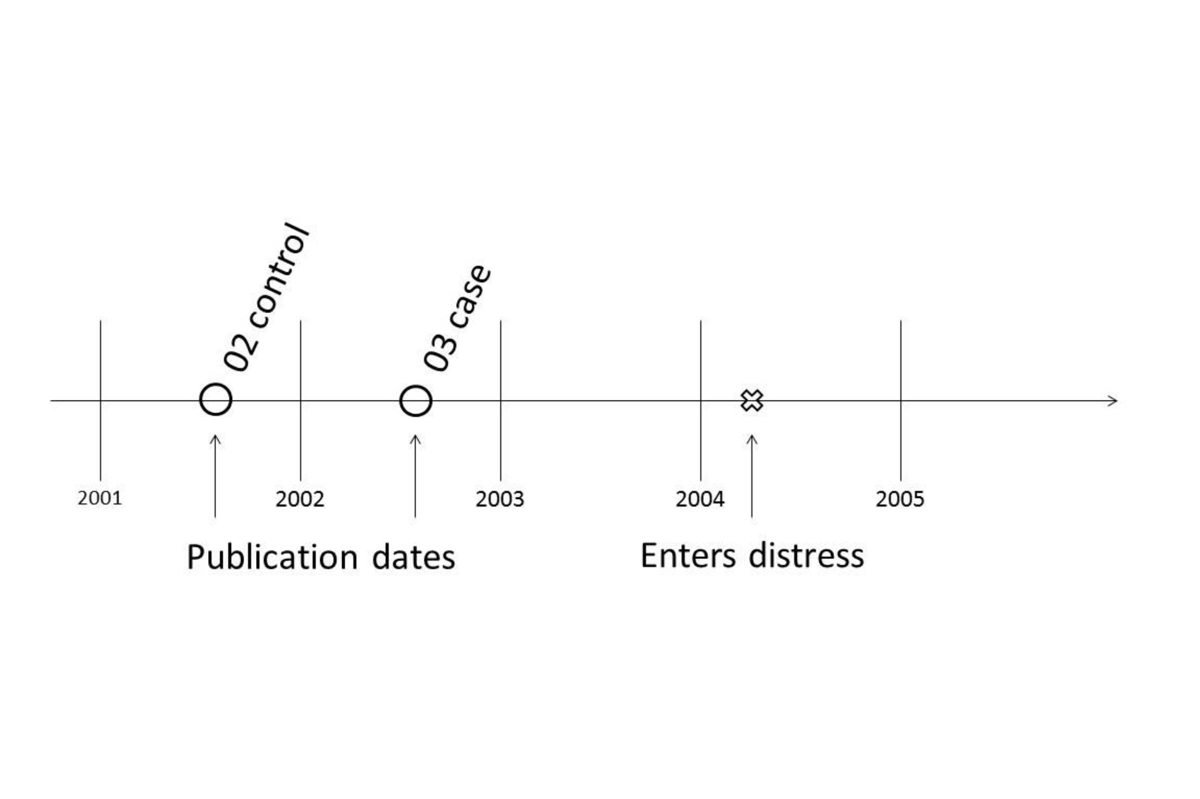
An event is entering into a distress period

within two years of last financial statement's publication date.

Some recover (recurrent events).

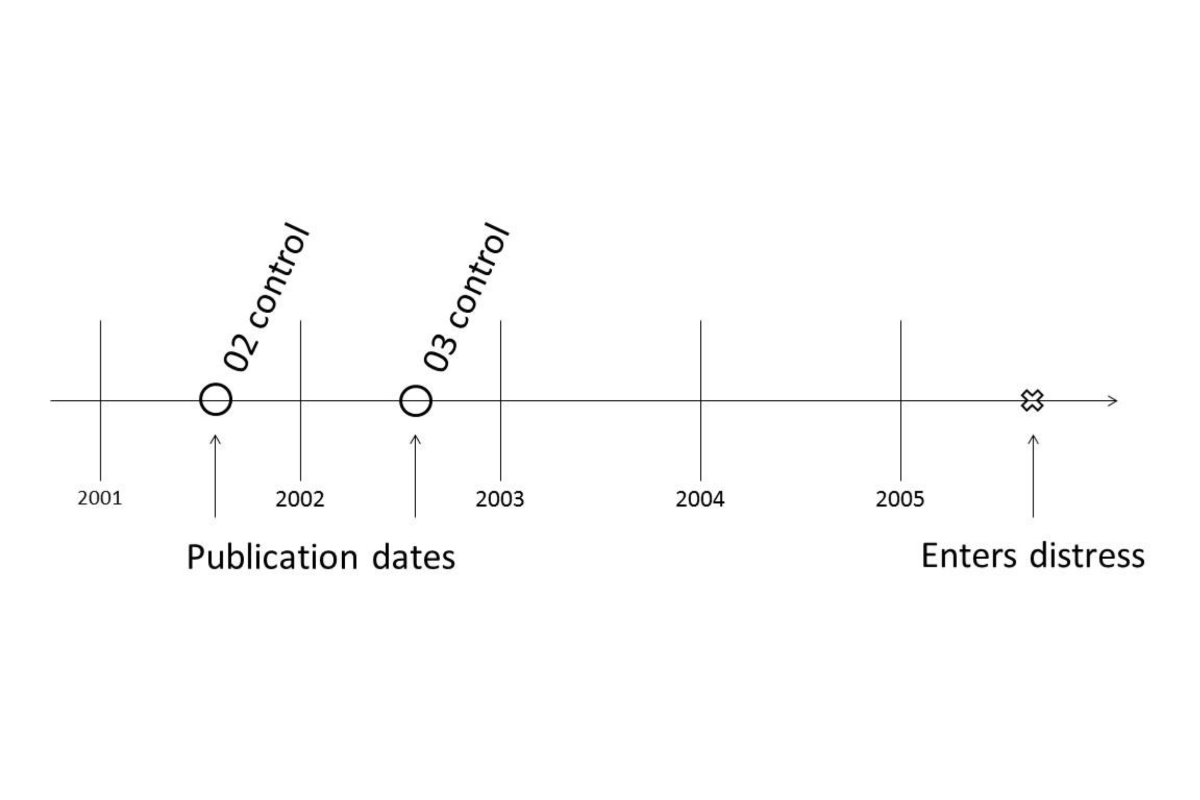
Some exit for other reasons (censoring).

Event example



Circle: a new financial statement. Cross: enters distress period.

Only control example



Generalized linear model (GLM)

Variable selection with two-stage LASSO procedure.

See Bühlmann and Van De Geer (2011) and Zhou (2010).

Examples of GLM PD models are Shumway (2001), Chava and Jarrow (2004), Beaver, McNichols, and Rhie (2005), and Campbell and Szilagyi (2008).

Model

$$E(Y_{it}|\vec{x}_{it}, \vec{z}_t) = p_{it}$$

$$\log(p_{it}/(1 - p_{it})) = \vec{\beta}^\top \vec{x}_{it} + \vec{\gamma}^\top \vec{z}_t$$

$$\underset{\vec{\beta}, \vec{\gamma}}{\operatorname{argmax}} \, L(\vec{Y}; \vec{\beta}, \vec{\gamma})$$

where $\vec{x}_{it}$ are firm specific variables and $\vec{z}_t$ are macro variables.

Final model has 57 firms specific variables and 1 macro variable.

Macro variables

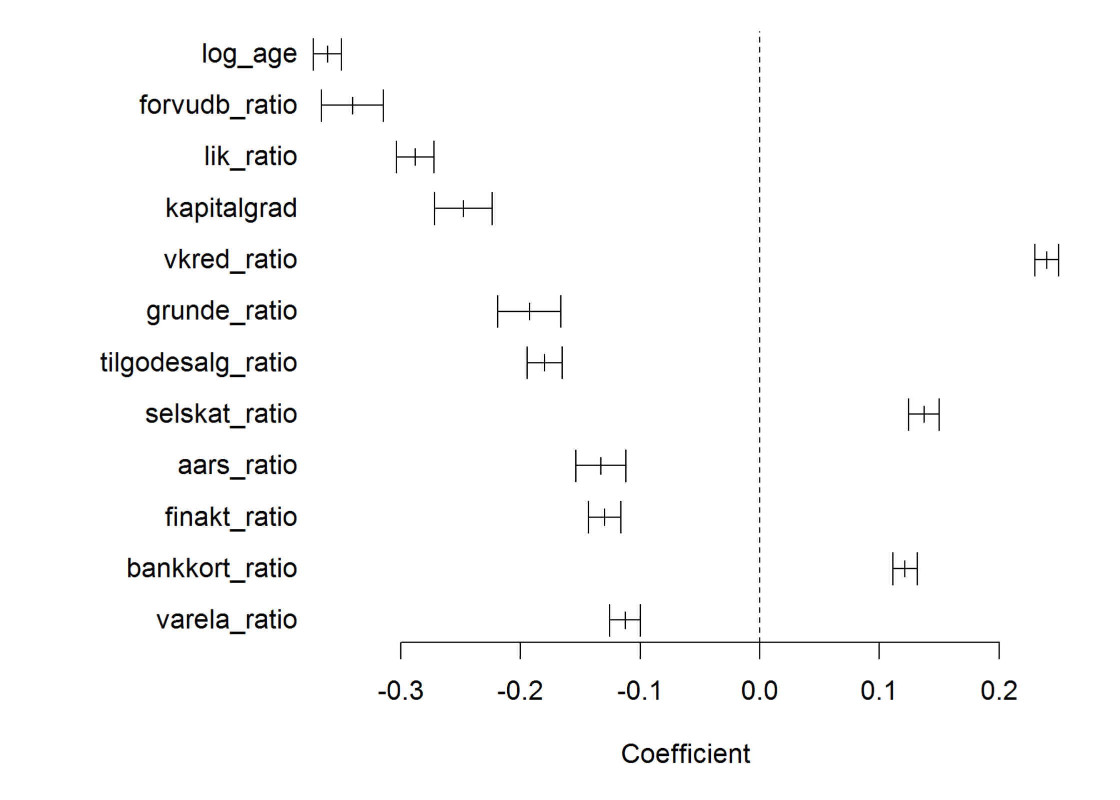
Take first component from PCA.

Explains 63.23pct. of the variation with 2015 data.

Variable	**Loading**
Unemployment rate	0.5636
Avg. bond yield less inflation	0.5752
Money market rate less inflation	0.5875
Inflation	0.0799

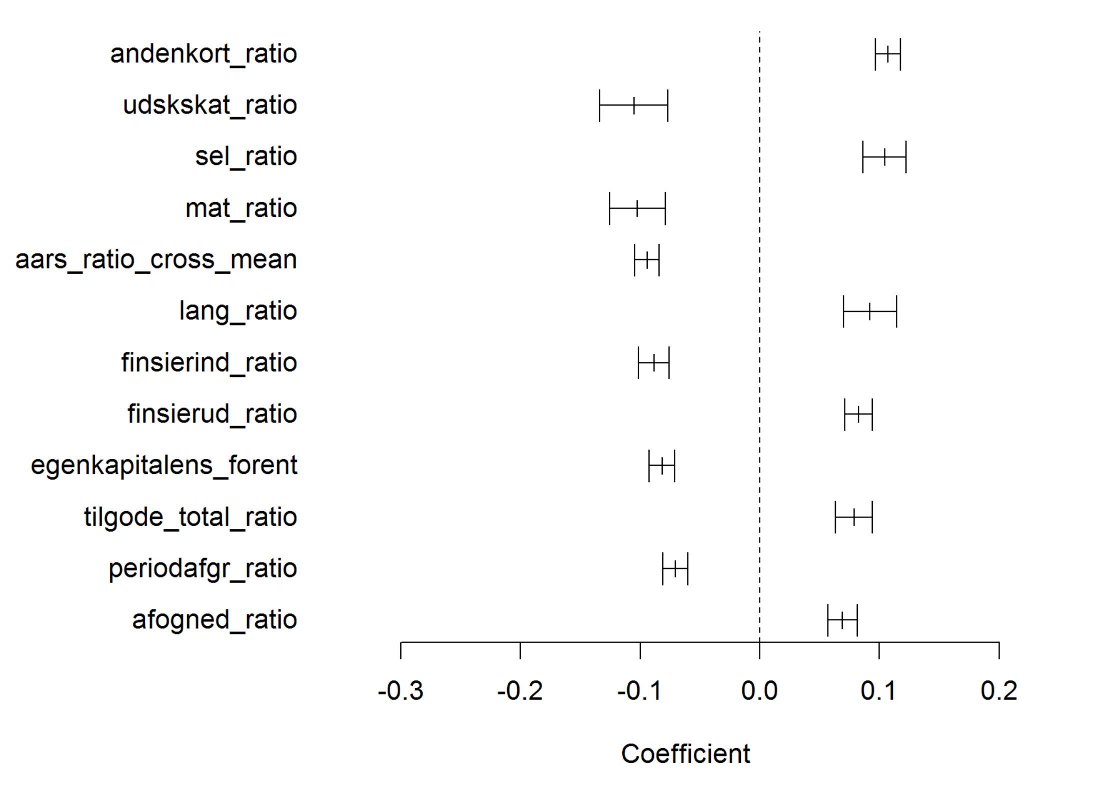
Use Q3 figures from previous year.

Estimates – standardized coefficients



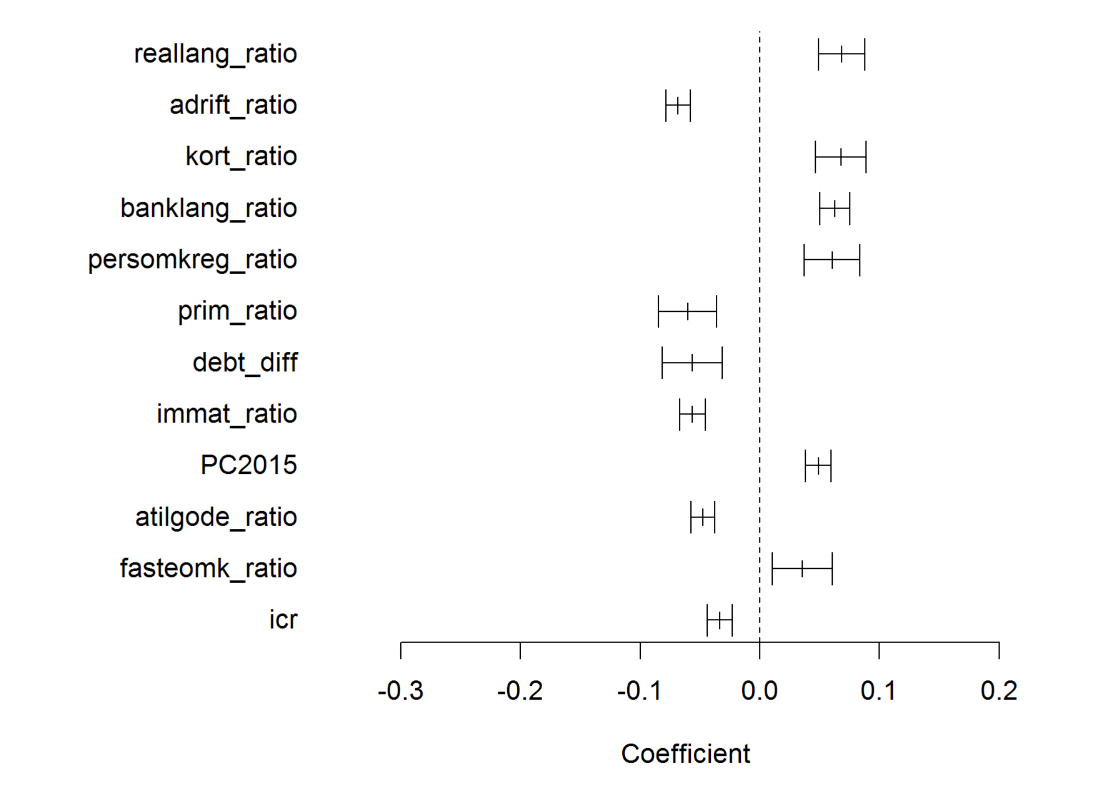
Inner line: coefficients with standardized covariates. Outer lines: 95% Wald confidence intervals.

Estimates – standardized coefficients



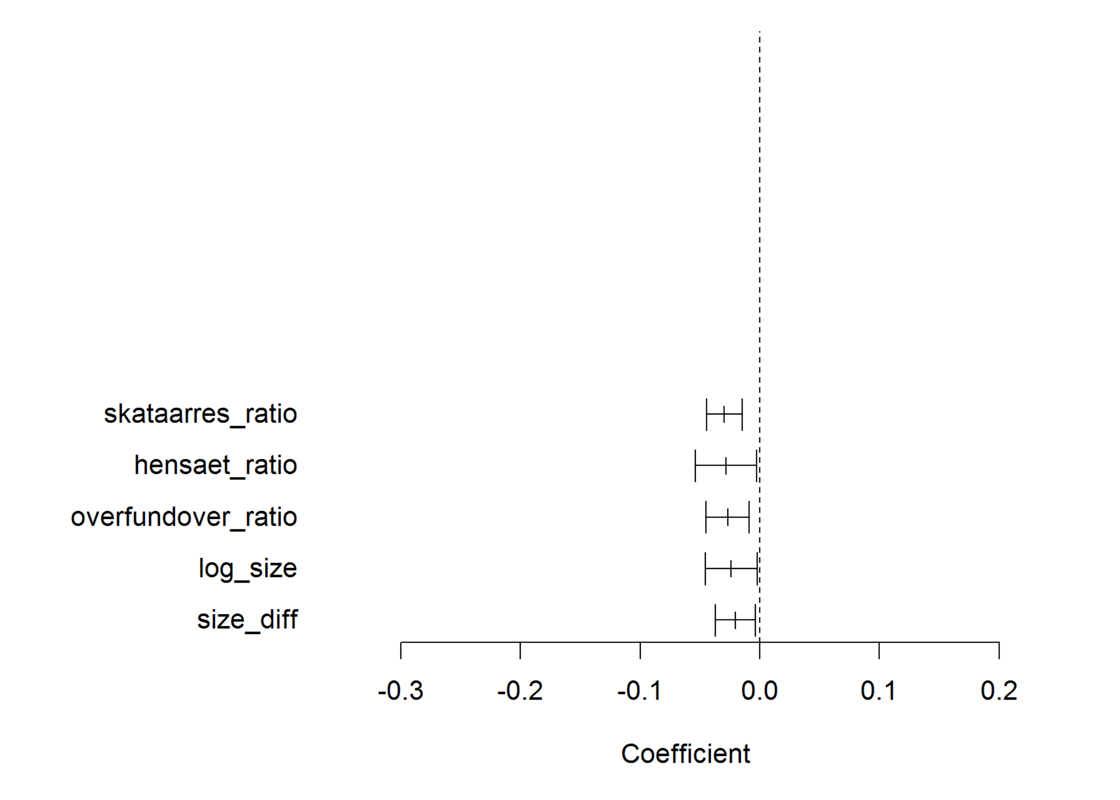
Inner line: coefficients with standardized covariates. Outer lines: 95% Wald confidence intervals.

Estimates – standardized coefficients



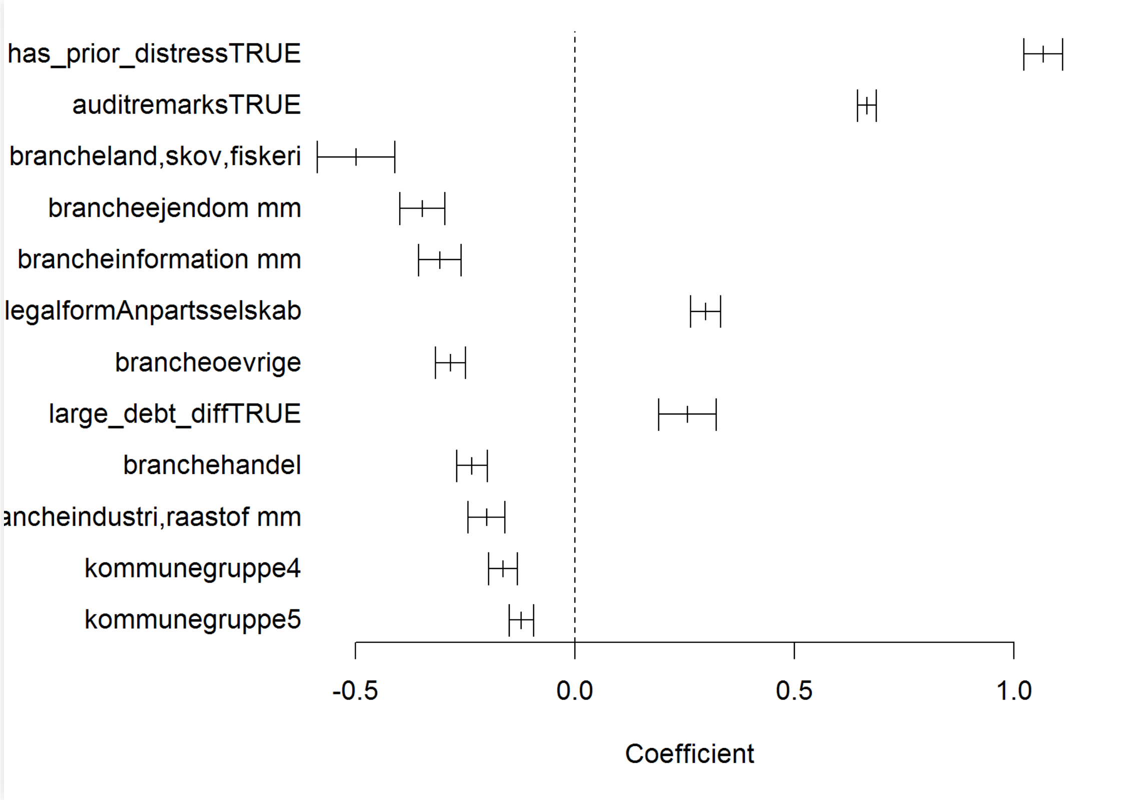
Inner line: coefficients with standardized covariates. Outer lines: 95% Wald confidence intervals.

Estimates – standardized coefficients



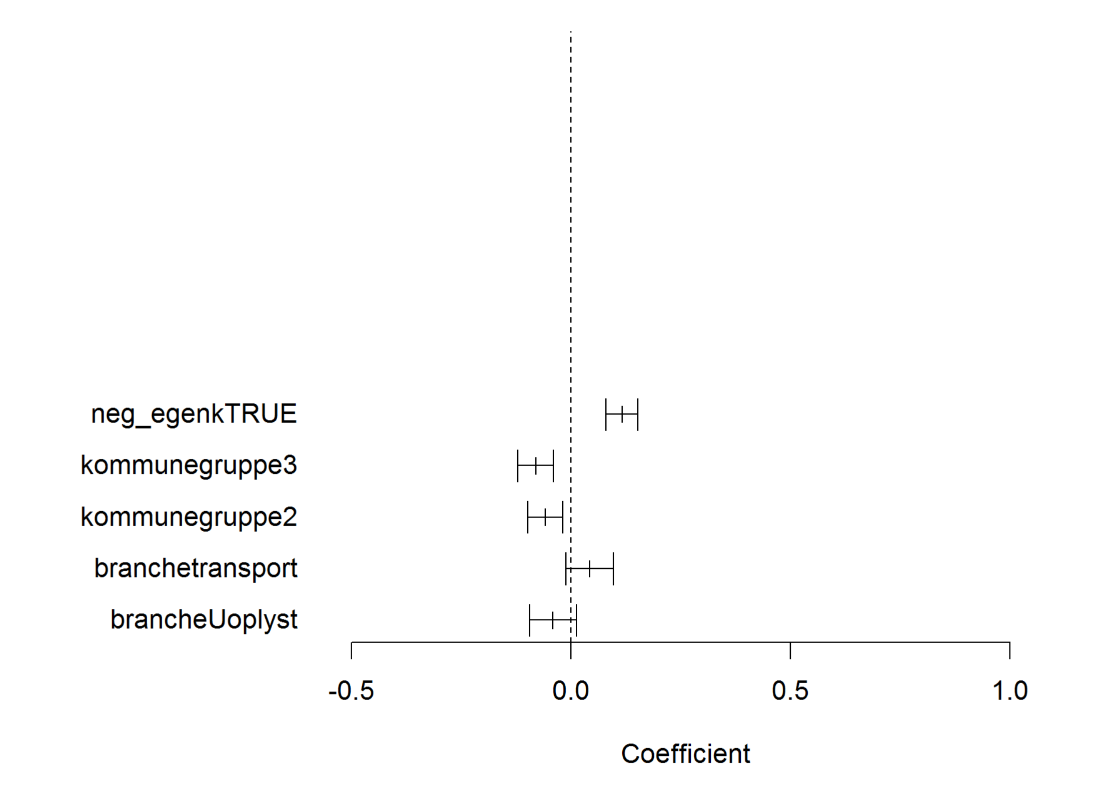
Inner line: coefficients with standardized covariates. Outer lines: 95% Wald confidence intervals.

Estimates – dummies



Inner line: coefficients. Outer lines: 95% Wald confidence intervals.

Estimates – dummies



Inner line: coefficients. Outer lines: 95% Wald confidence intervals.

Generalized additive model (GAM)

Model non-linear effects.

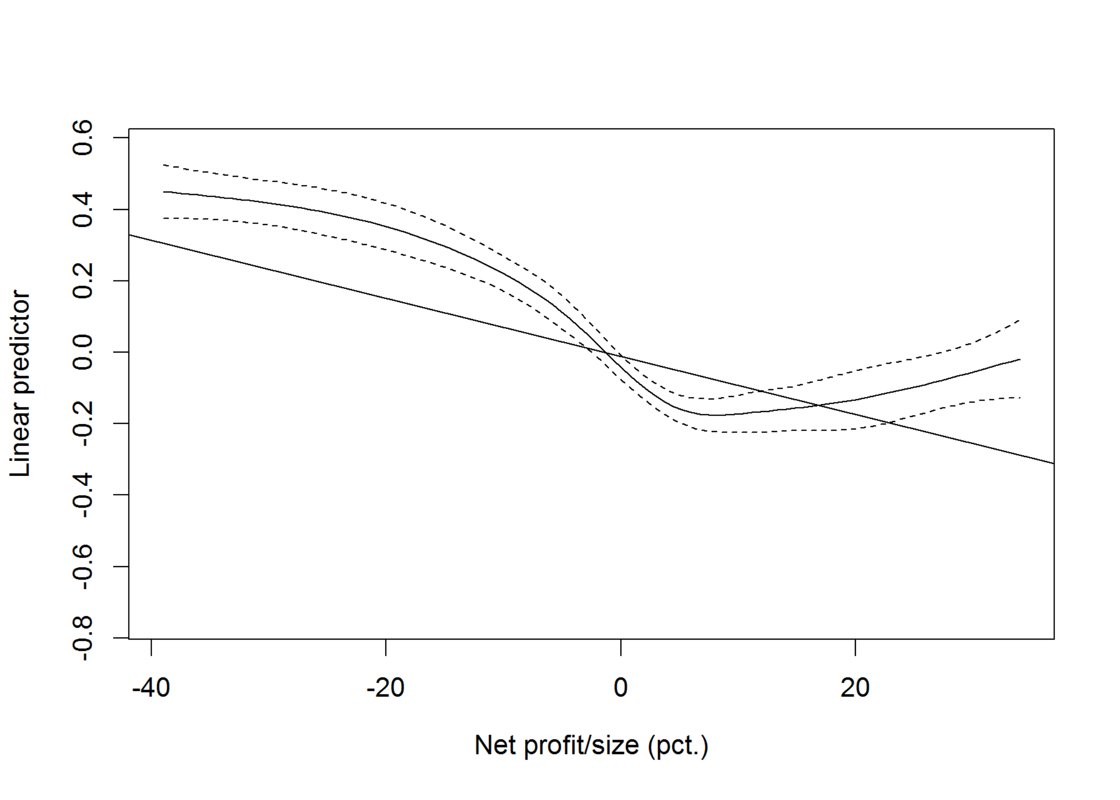
Model	**df**	**AIC**
GLM	59	$327\,851$
GAM	197.2	$323\,895$

More complex and much better fit.

Examples of GAM PD models are Dakovic, Czado, and Berg (2010) and Berg (2007).

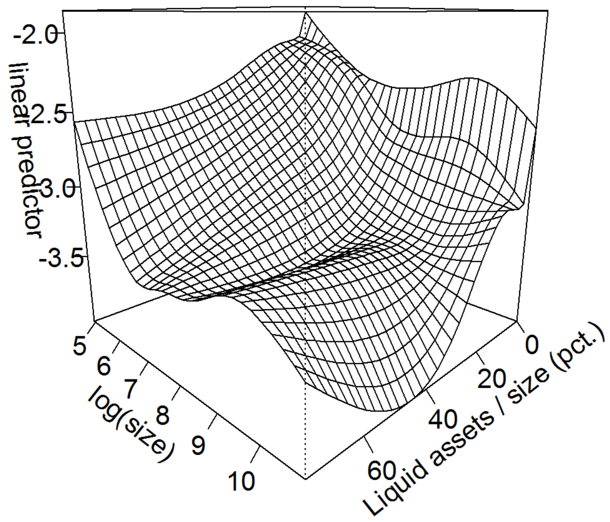
Cubic regression splines with second order penalties.
See S. Wood (2017) and S. N. Wood, Goude, and Shaw (2015).

Marginal effect in GLM and GAM

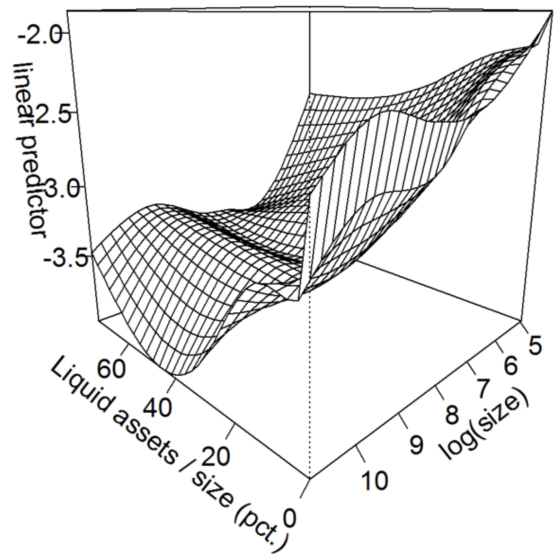


Line: effect in GLM. Curve: effect in GAM with 2 standard error bounds.
Higher implies higher distress probability.

Marginal effect GAM



Marginal effect GAM



Pros and cons

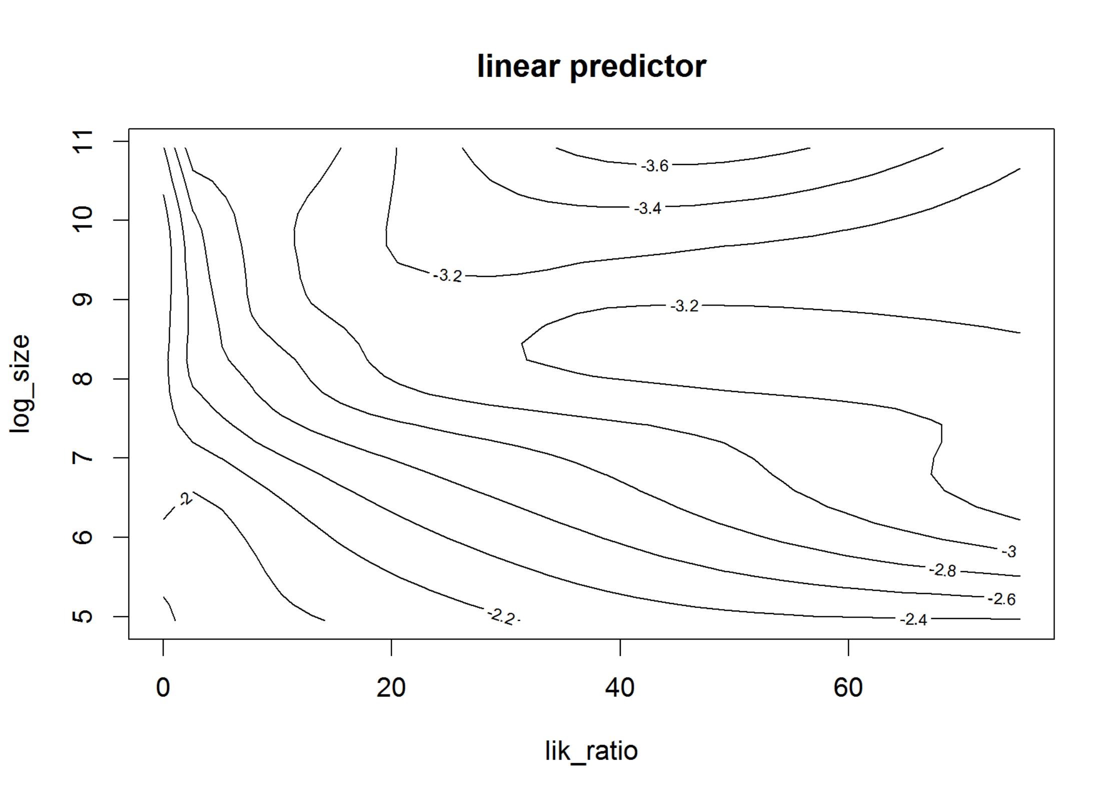
Pros

- Fit more complex models.
- Control complexity of each covariate
- Easy to judge marginal effects and interactions.

Cons

- Have to specify effect of each covariate (slope, spline or interactions).

Marginal effect GAM



Higher implies higher distress probability.

Model

$$
\begin{aligned}
E(Y_{it}|\vec{x}_{it}, \vec{z}_t) &= p_{it} \\
\log(p_{it}/(1-p_{it})) &= \vec{\beta}_1^\top \vec{x}_{1it} + \sum_{j=1}^{q} \vec{\beta}_{2j}^\top \vec{f}_j(x_{2ijt}) \\
&\quad + \sum_{k=1}^{r} \sum_{(j,l,s)\in S_k} \vec{\beta}_{3kj} f_{ls}(x_{3ilt}) \vec{f}_j(x_{3ijt}) \\
&\quad + \vec{\gamma}^\top \vec{z}_t \\
\operatorname*{argmax}_{\vec{\beta},\vec{\gamma}} & L(\vec{Y}; \vec{\beta}, \vec{\gamma}) - \vec{\beta} S_{\vec{\lambda}} \vec{\beta}
\end{aligned}
$$

Use generalized cross validation to select $\vec{\lambda}$.

Gradient tree boosting (GB)

Used implementation is common method in winning solutions on kaggle.com.

kaggle

dmlc XGBoost

Examples of GB PD models are
Alfaro et al. (2008), Zięba, Tomczak, and Tomczak (2016) and Jones, Johnstone, and Wilson (2017).

Sequential approximations of the gradient.
See Friedman (2001), Friedman (2002), and Bühlmann and Hothorn (2007).

Pros and cons

Pros

- Fit (very) complex models.
- Automated / easy to use.

Cons

- Limited control of complexity of each covariate.
- Hard to judge marginal effects.

Estimation

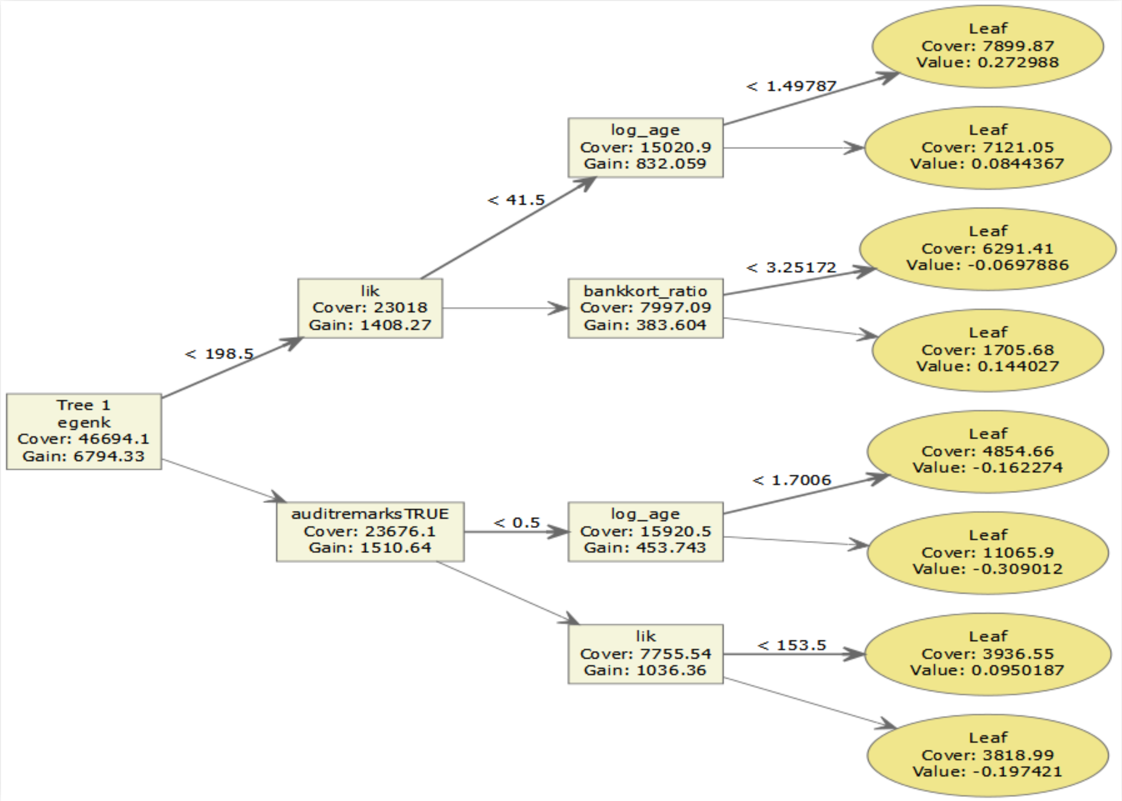
Initialize linear predictors by $\theta_{it} = \text{logit}(\bar{p})$

Define the loss function, L, by

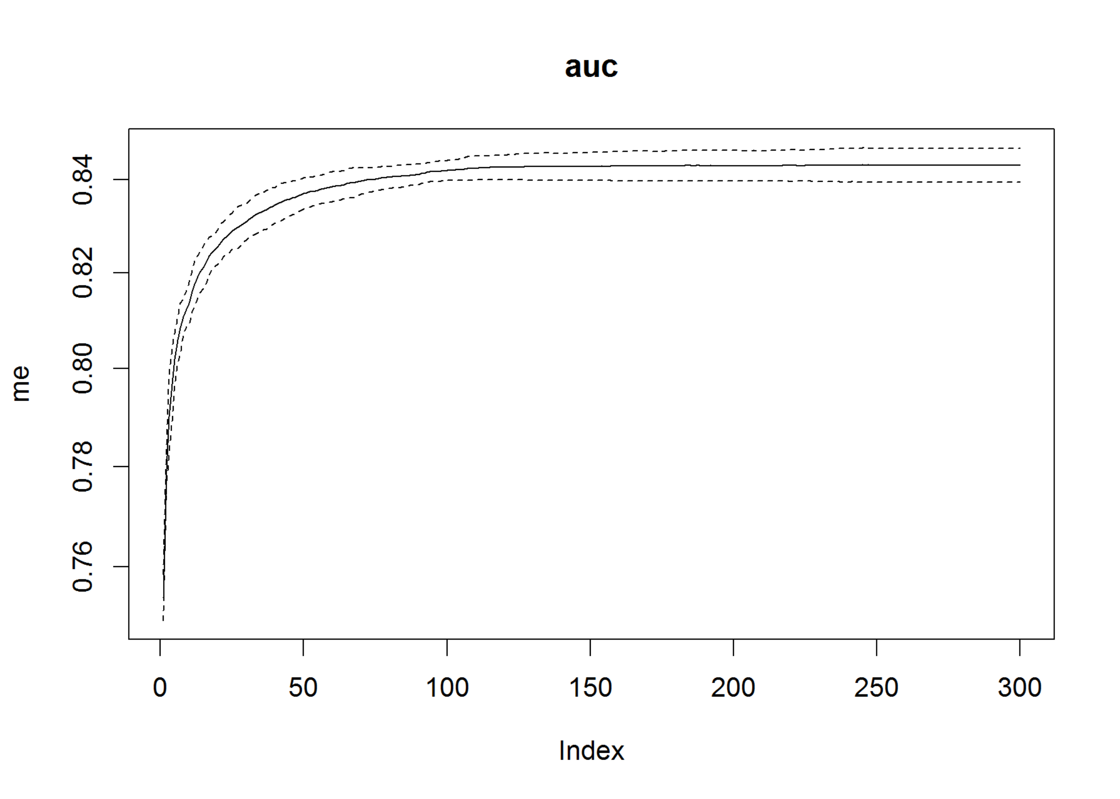
$$L(\theta; y) \;= y_{it} \log p(\theta) + (1 - y_{it}) \log(1 - p(\theta)), \qquad p(\theta) = \text{logit}^{-1}(\theta)$$
$$L = \sum_{t=1}^{d} \sum_{i \in R_t} L(\theta_{it}; y_{it})$$

1. Compute gradient of loss function, $\sum_{t=1}^{d} \sum_{i \in R_t} L'(\theta_{it}; y_{it})$
2. Train simple model (shallow tress) to fit the gradient using x_{it} and z_t .
3. Update linear predictors, θ_{it}s, and repeat 1.

Example with first tree

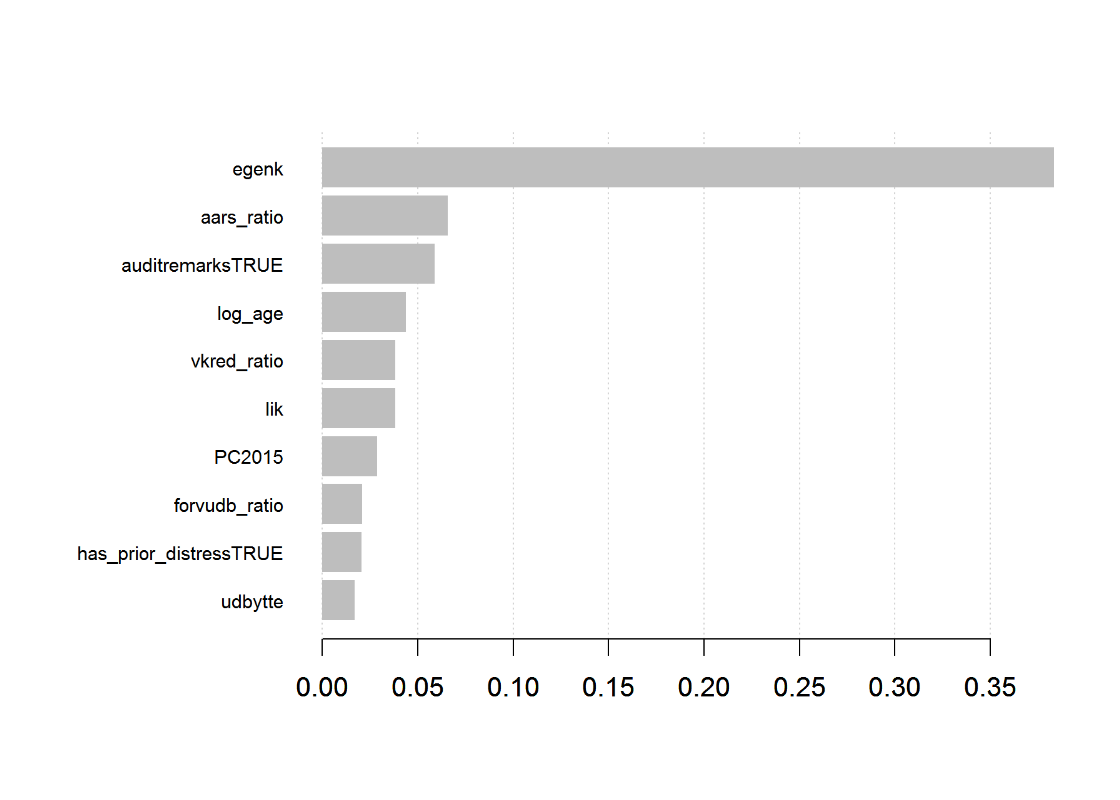


Choosing shrinkage parameter and depth



Select tree depth and number of boosting iteration with K-fold cross validation.

Variable importance



Relative improvement in log likelihood. See Friedman (2001).

Comparison

Out sample test using expanding window.

E.g., estimate with 2001-2007 data and forecast 2009.

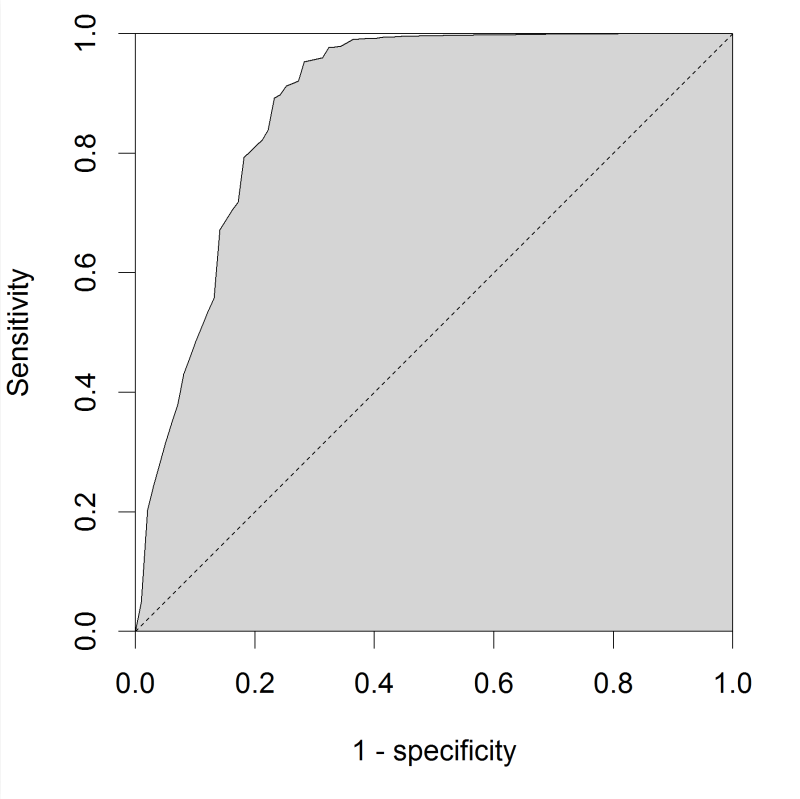
What is the AUC

Probability of **ranking** random distressed firm at higher risk than random non-distressed firm.

0.5 is random guessing and 1 is perfect order.

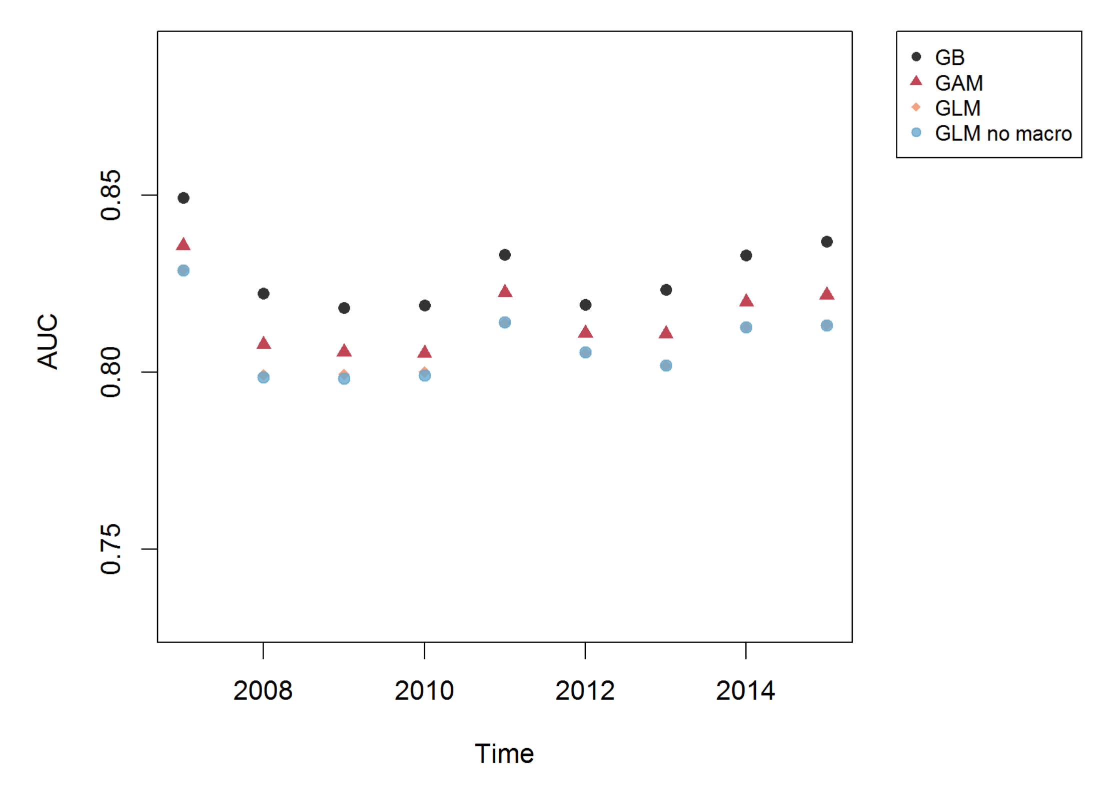
$$\text{AUC} = \frac{1}{2}(\text{Gini coef} - 1)$$

What is the AUC



AUC: Area Under the ROC Curve.

AUC – all firms



Larger is better.

Conclussion

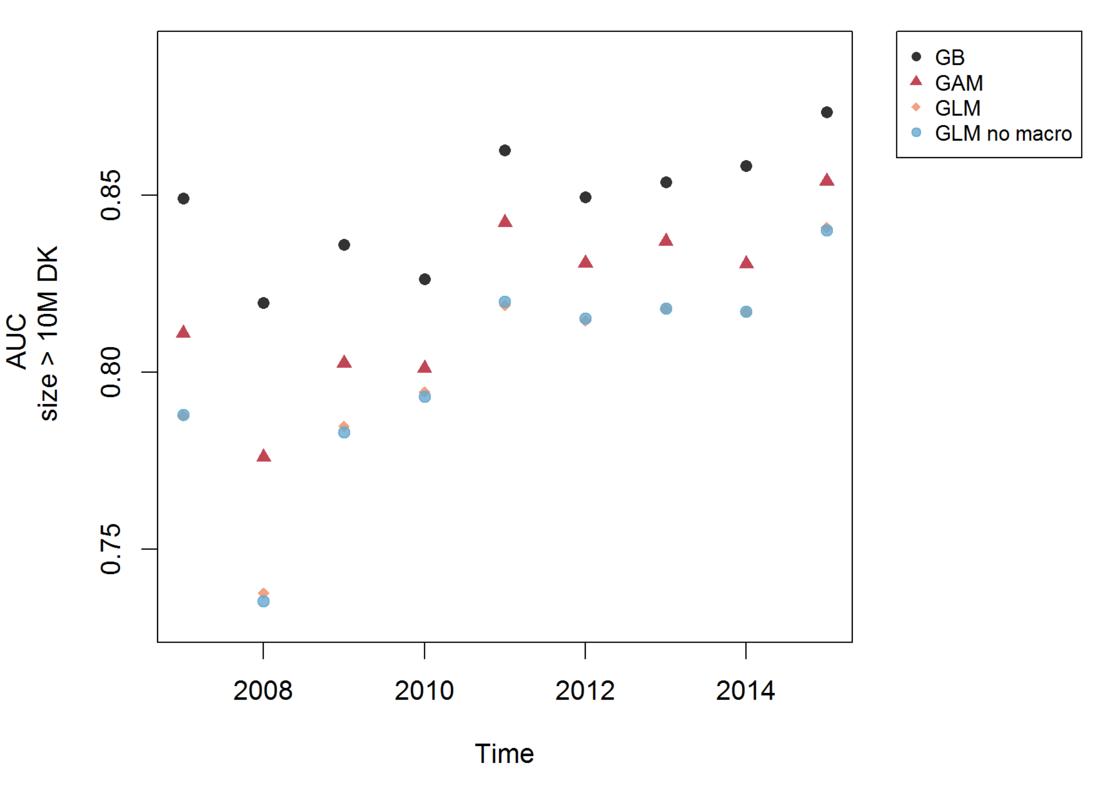
GB is better at **ranking.**

Roughly 0.01 higher AUC.

This is far from the 0.10 to 0.30 differences shown recently

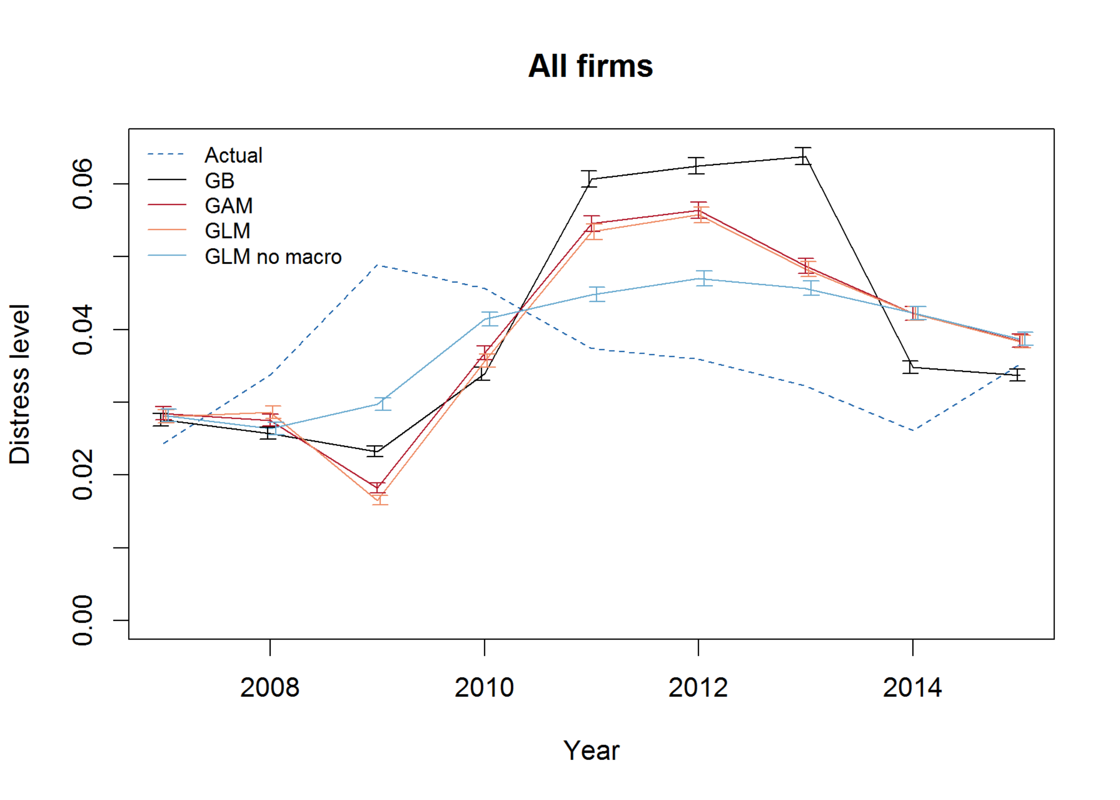
in Zięba, Tomczak, and Tomczak (2016) and Jones, Johnstone, and Wilson (2017).

AUC – large firms



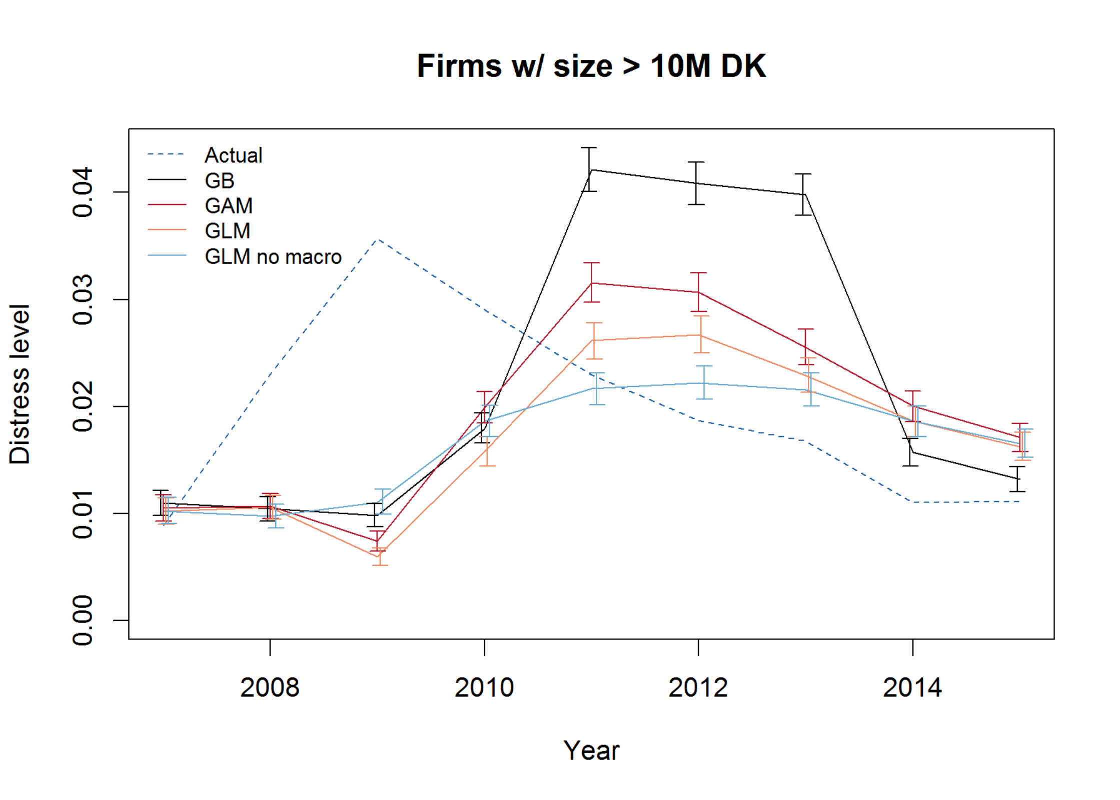
21% of the firms and 92% of total short and long term debt.

Predicted level of distress – all



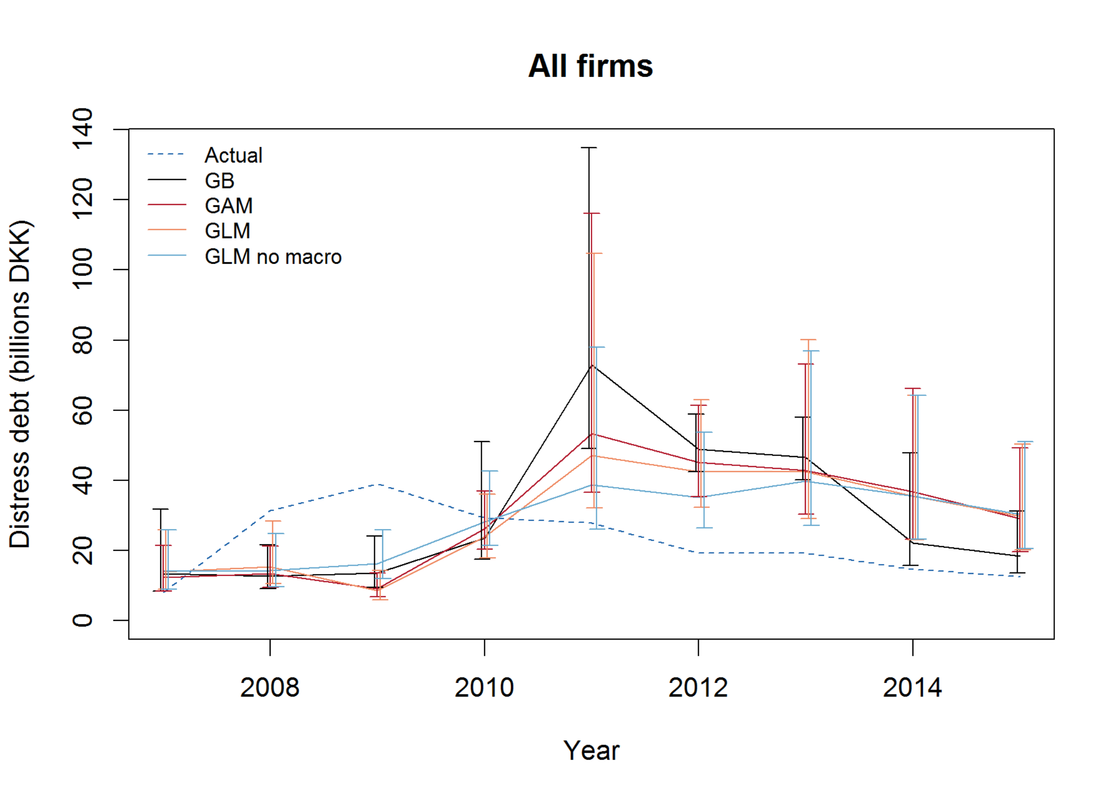
Bars are simulated 5 pct. and 95 pct. quantiles.

Predicted level of distress – large



Bars are simulated 5 pct. and 95 pct. quantiles.

Risk weighted debt



Debt is long and short term debt on balance sheet.

Conclussion

All models fail to fit the distress level.

They fail to capture temporal changes in the level.

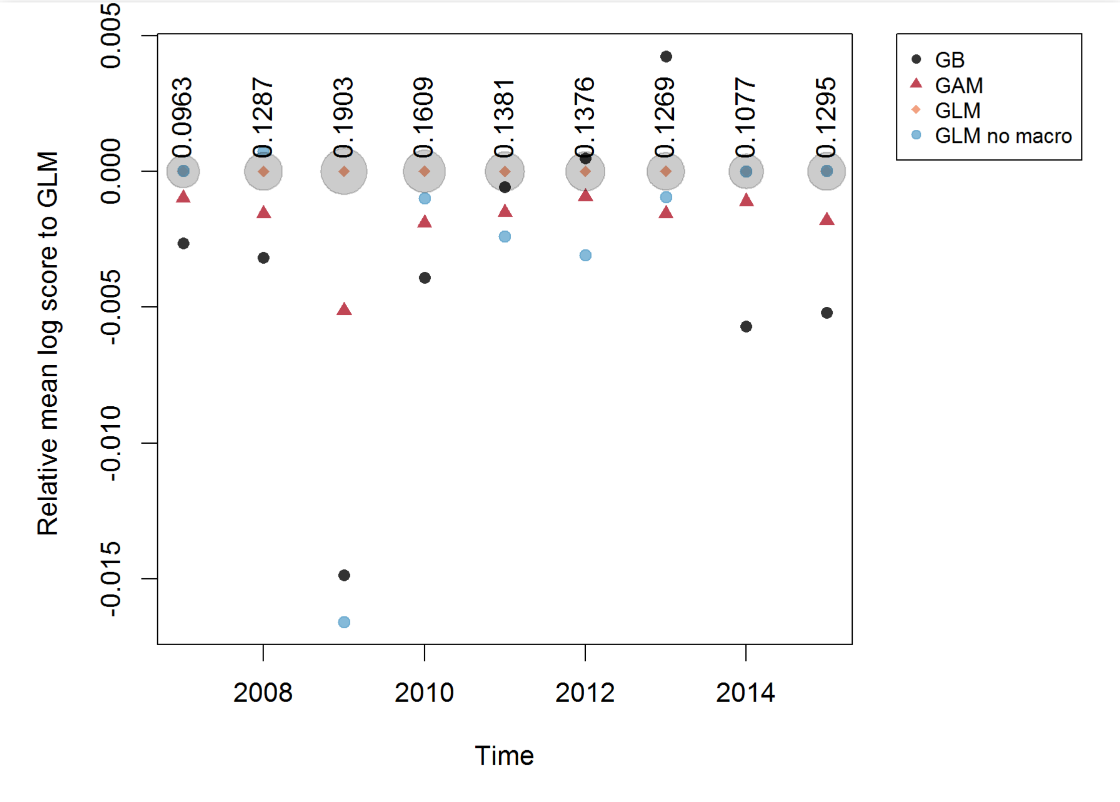
Particular GB model performs poorly in some periods.

The forecasted 5% to 95% intervals are too narrow.

Mean log score

$$\text{avg log score}_t = -\frac{1}{|R_t|}\sum_{i \in R_t}\left(y_{it}\log \hat{p}_{it} + (1 - y_{it})\log(1 - \hat{p}_{it})\right)$$

Mean log score – all firms



Y-value is relative to the GLM. Numbers indicate the mean log score for the GLM. Circle areas are proportional to mean log score of GLM. Smaller is better.

Generalized linear mixed models (GLMM)

Add time dependent random effects.

See D. M. Bates and DebRoy (2004) and D. Bates et al. (2015).

Examples of GLMM PD models are Duffie et al. (2009), Koopman, Lucas, and Schwaab (2011), Azizpour, Giesecke, and Schwenkler (2016), Nickerson and Griffin (2017), and Kwon and Lee (2018).

Model

$$E(Y_{it}|\vec{x}_{it}, \vec{z}_t, \vec{u}_{it}, \vec{\epsilon}_t) = p_{it}$$

$$\log(p_{it}/(1-p_{it})) = \vec{\beta}^\top \vec{x}_{it} + \vec{\gamma}^\top \vec{z}_t + \vec{\epsilon}_t^\top \vec{u}_{it} \qquad \vec{\epsilon}_t \sim N(\vec{0}, Q)$$

$$\underset{\vec{\beta}, \vec{\gamma}, Q}{\operatorname{argmax}} L(\vec{Y}; \vec{\beta}, \vec{\gamma}, Q)$$

where $\vec{u}_{it}$ are covariates for the random effects.

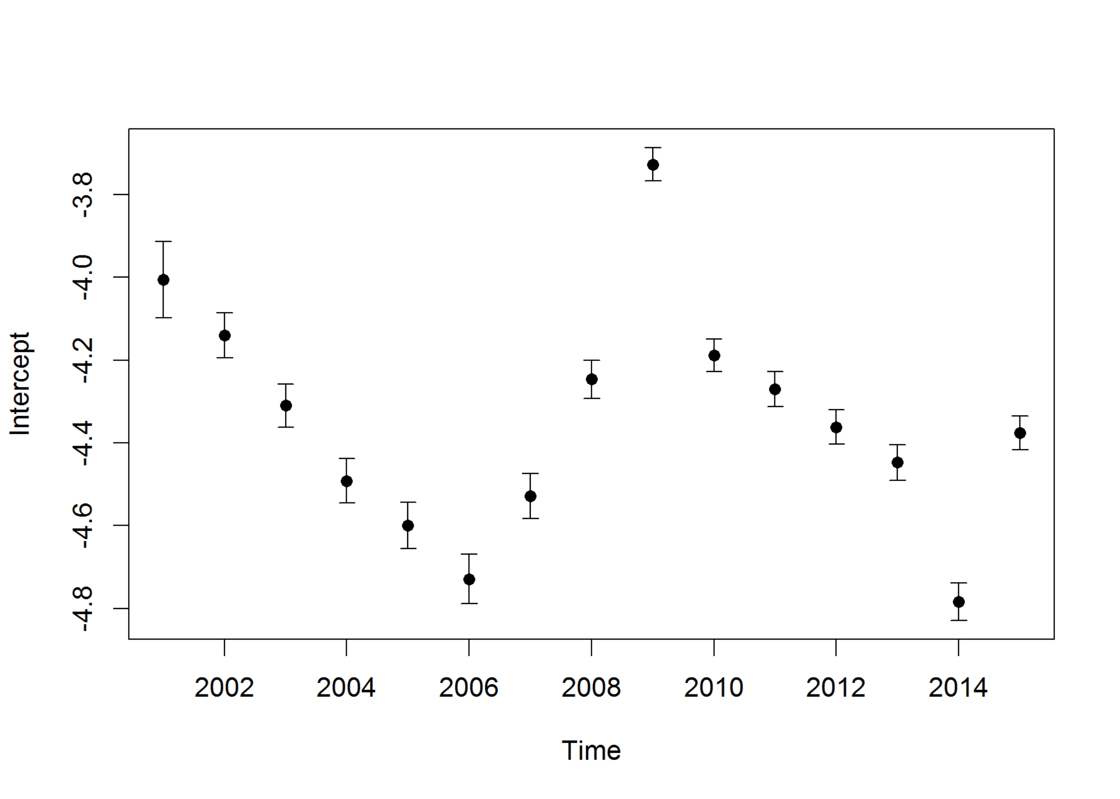
Random intercept

$$\beta_{0t} = \beta_0 + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma^2)$$

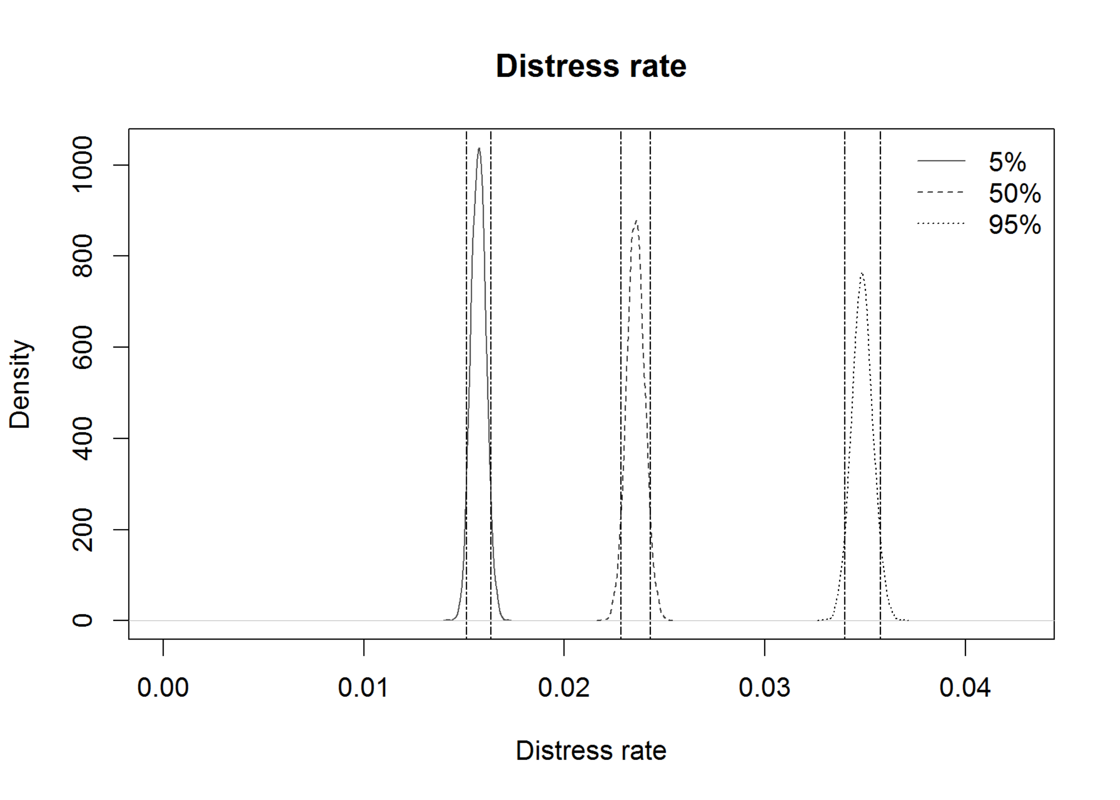
$$\hat{\sigma} = 0.266$$

Conservative $\chi^2_{(1)}$ statistics is 1590 with $H_0 : \sigma = 0$

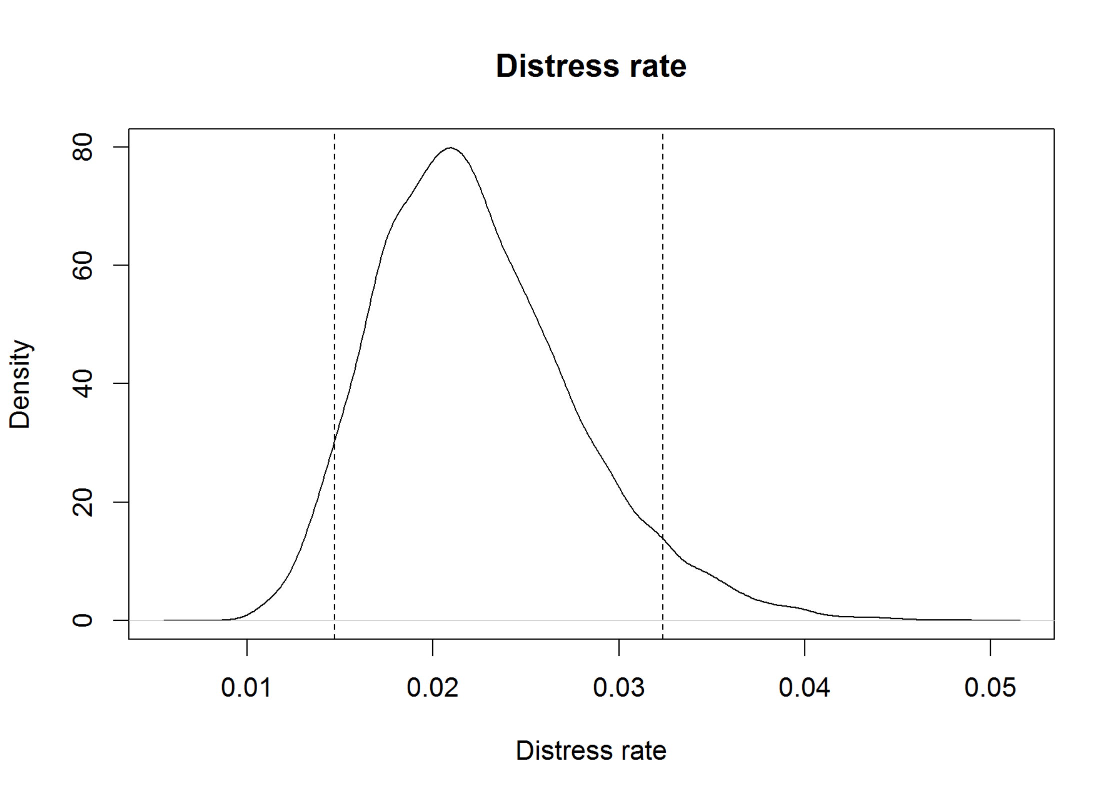
Conditional modes



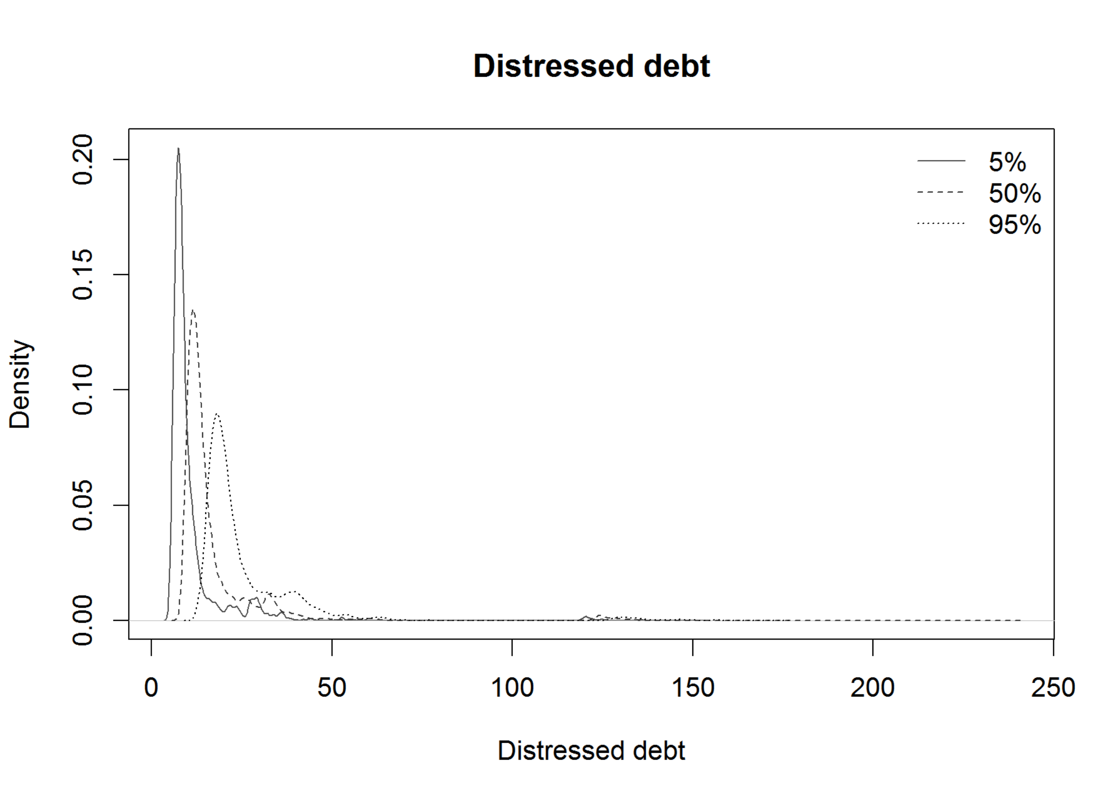
2017 prediction – distresses



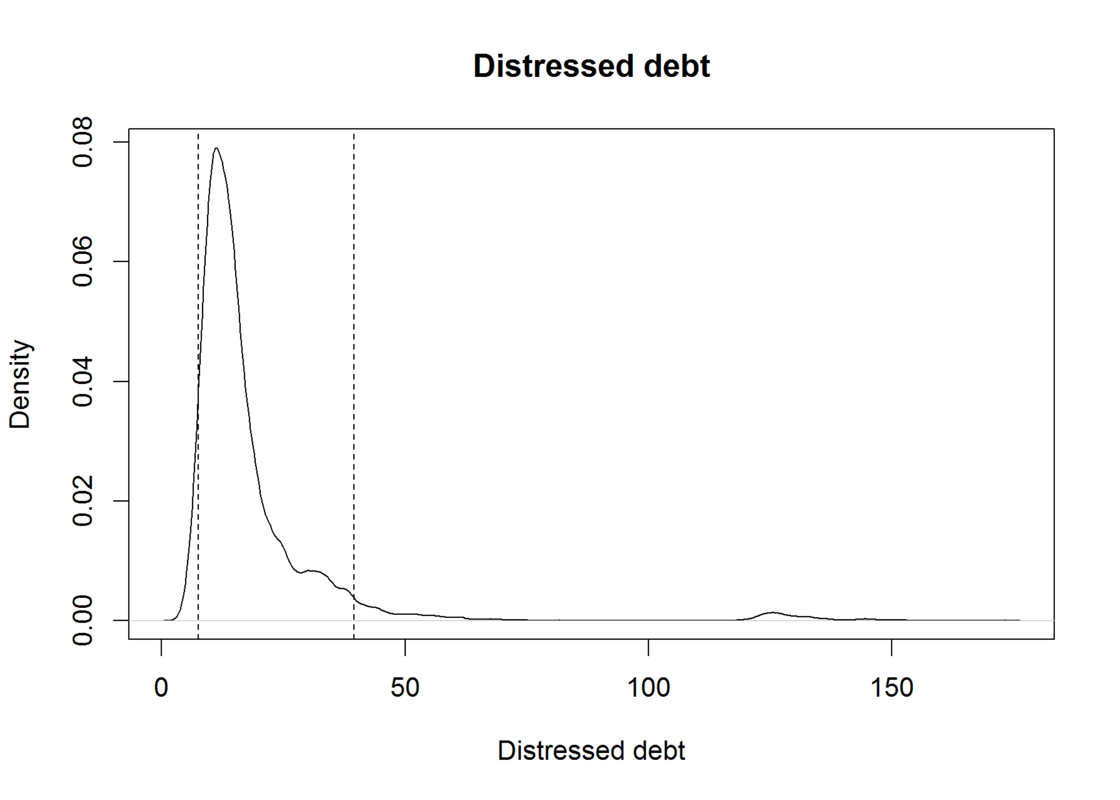
2017 prediction – distresses



2017 prediction – debt



2017 prediction – debt



Mean is 18.5 billions DK.

Conclussion

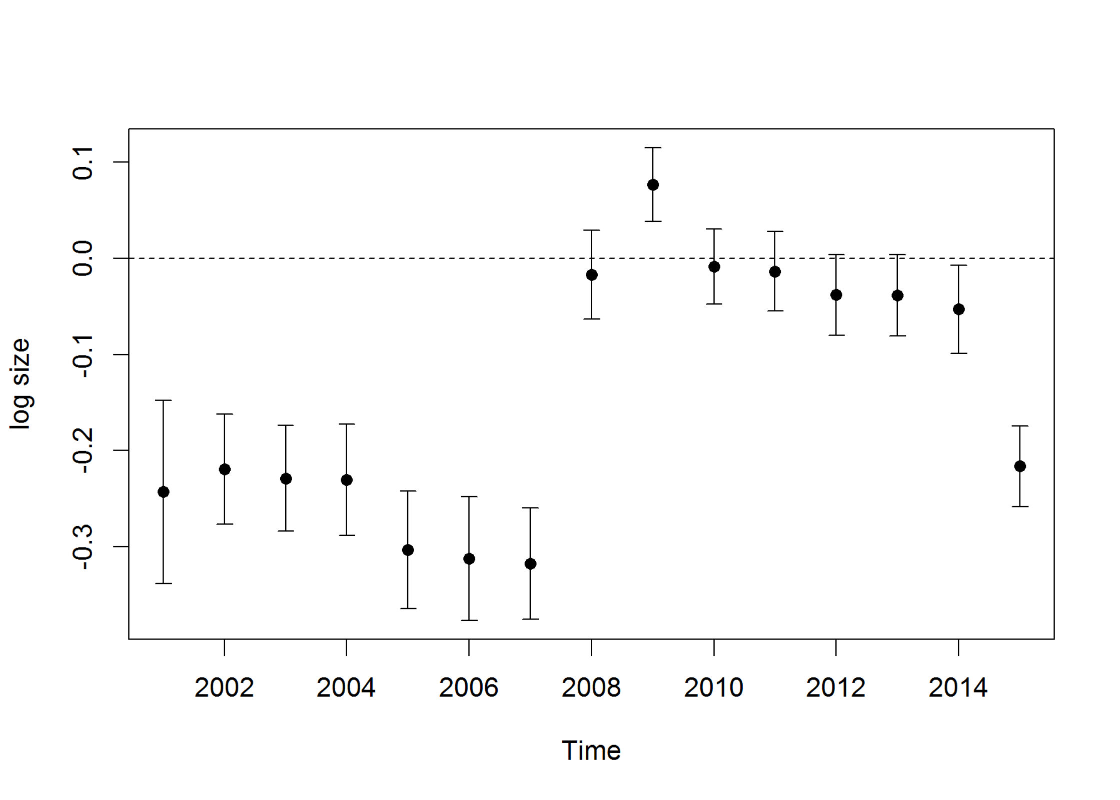
GLMMs let's us model the uncertainty.

Key when evaluating riskiness of a portfolio of firms and care about tail events.

Random intercept and log size slope

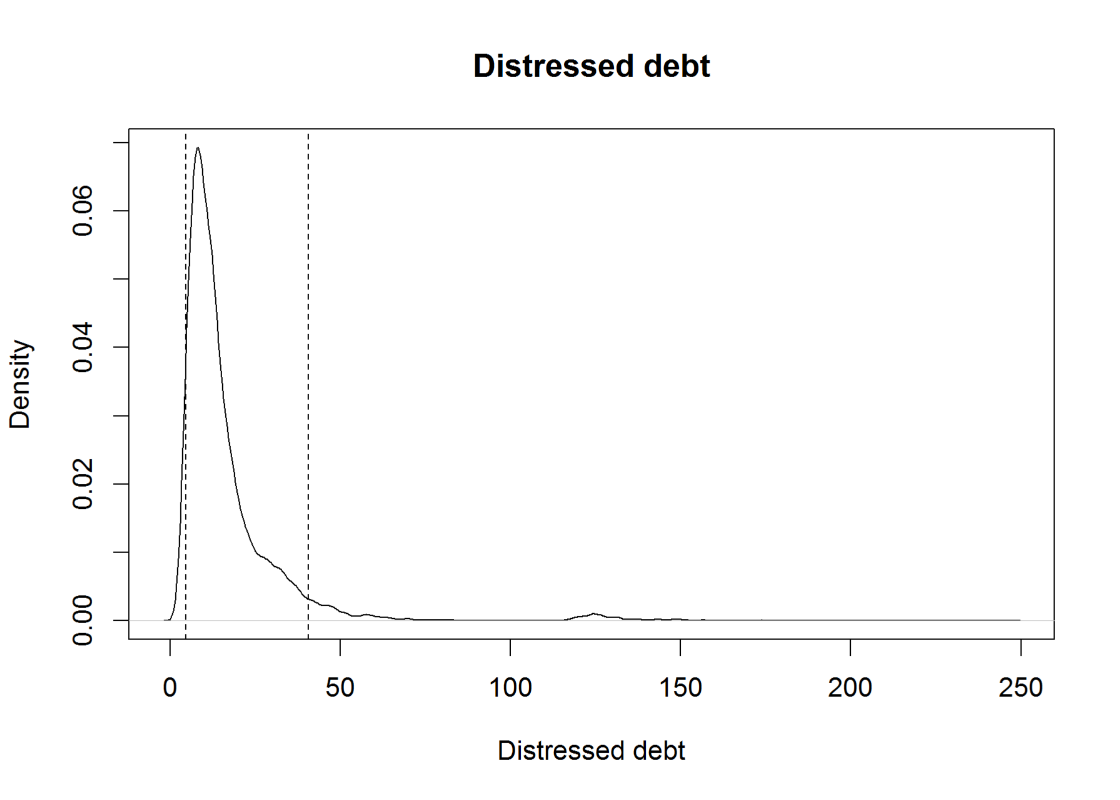
$$\begin{pmatrix} \epsilon_{\text{intercept},t} \\ \epsilon_{\log(\text{size}),t} \end{pmatrix} \sim N\left(\vec{0}, Q\right), \qquad \widehat{Q} = \begin{pmatrix} 0.0965 & 0.0291 \\ 0.0291 & 0.0175 \end{pmatrix}$$

Random intercept and log size slope



Fixed slope estimate is -0.1443

2017 prediction – debt



Mean is 16.4 billions DK.

Conclusions

There are gains of more complex models in terms of AUC.
A ranking measure.

The gains in AUC are smaller than suggested in recent papers.

The more complex models does not overcome issues with temporal effects.

The temporal effects can be addressed with random effects models.

References

Alfaro, Esteban, Noelia García, Matías Gámez, and David Elizondo. 2008. "Bankruptcy Forecasting: An Empirical Comparison of Adaboost and Neural Networks." *Decision Support Systems* 45 (1): 110–22. doi:https://doi.org/10.1016/j.dss.2007.12.002.

Azizpour, Shahriar, Kay Giesecke, and Gustavo Schwenkler. 2016. "Exploring the Sources of Default Clustering."

Bates, Douglas M, and Saikat DebRoy. 2004. "Linear Mixed Models and Penalized Least Squares." *Journal of Multivariate Analysis* 91 (1): 1–17. doi:https://doi.org/10.1016/j.jmva.2004.04.013.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software, Articles* 67 (1): 1–48. doi:10.18637/jss.v067.i01.

Beaver, William H., Maureen F. McNichols, and Jung-Wu Rhie. 2005. "Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy." *Review of Accounting Studies* 10 (1): 93–122. doi:10.1007/s11142-004-6341-9.

Berg, Daniel. 2007. "Bankruptcy Prediction by Generalized Additive Models." *Applied Stochastic Models in Business and Industry* 23 (2). John Wiley & Sons, Ltd.: 129–43. doi:10.1002/asmb.658.

Bühlmann, Peter, and Torsten Hothorn. 2007. "Boosting Algorithms: Regularization, Prediction and Model Fitting." *Statistical Science* 22 (4). Institute of Mathematical Statistics: 477–505. http://www.jstor.org/stable/27645854.

Bühlmann, Peter, and Sara Van De Geer. 2011. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Science & Business Media.

Campbell, Jens, John Y. And Hilscher, and Jan Szilagyi. 2008. "In Search of Distress Risk." *The Journal of Finance* 63 (6). Blackwell Publishing Inc: 2899–2939. doi:10.1111/j.1540-6261.2008.01416.x.

Chava, Sudheer, and Robert A. Jarrow. 2004. "Bankruptcy Prediction with Industry Effects *." *Review of Finance* 8 (4): 537–69. doi:10.1093/rof/8.4.537.

Dakovic, Rada, Claudia Czado, and Daniel Berg. 2010. "Bankruptcy Prediction in Norway: A Comparison Study." *Applied Economics Letters* 17 (17). Routledge: 1739–46. doi:10.1080/13504850903299594.

Duffie, Darrell, Andreas Eckner, Guillaume Horel, and Leandro Saita. 2009. "Frailty Correlated Default." *The Journal of Finance* 64 (5). Blackwell Publishing Inc: 2089–2123. doi:10.1111/j.1540-6261.2009.01495.x.

Friedman, Jerome H. 2001. "Greedy Function Approximation: A Gradient Boosting Machine." *The Annals of Statistics* 29 (5). Institute of Mathematical Statistics: 1189–1232. http://www.jstor.org/stable/2699986.

———. 2002. "Stochastic Gradient Boosting." *Computational Statistics & Data Analysis* 38 (4): 367–78. doi:https://doi.org/10.1016/S0167-9473(01)00065-2.

Jones, Stewart, David Johnstone, and Roy Wilson. 2017. "Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks." *Journal of Business Finance & Accounting* 44 (1-2): 3–34. doi:10.1111/jbfa.12218.

Koopman, Siem Jan, André Lucas, and Bernd Schwaab. 2011. "Modeling Frailty-Correlated Defaults Using Many Macroeconomic Covariates." *Journal of Econometrics* 162 (2): 312–25. doi:https://doi.org/10.1016/j.jeconom.2011.02.003.

Kwon, Tae Yeon, and Yoonjung Lee. 2018. "Industry Specific Defaults." *Journal of Empirical Finance* 45: 45–58. doi:https://doi.org/10.1016/j.jempfin.2017.10.002.

Nickerson, Jordan, and John M. Griffin. 2017. "Debt Correlations in the Wake of the Financial Crisis: What Are Appropriate Default Correlations for Structured Products?" *Journal of Financial Economics* 125 (3): 454–74. doi:https://doi.org/10.1016/j.jfineco.2017.06.011.

Shumway, Tyler. 2001. "Forecasting Bankruptcy More Accurately: A Simple Hazard Model." *The Journal of Business* 74 (1). The University of Chicago Press: 101–24. http://www.jstor.org/stable/10.1086/209665.

Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman; Hall/CRC.

Wood, Simon N., Yannig Goude, and Simon Shaw. 2015. "Generalized Additive Models for Large Data Sets." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64 (1): 139–55. doi:10.1111/rssc.12068.

Zhou, S. 2010. "Thresholded Lasso for high dimensional variable selection and statistical estimation." *ArXiv E-Prints*, February.

Zięba, Maciej, Sebastian K. Tomczak, and Jakub M. Tomczak. 2016. "Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction." *Expert Systems with Applications* 58: 93–101. doi:https://doi.org/10.1016/j.eswa.2016.04.001.