



Technische  
Universität  
Braunschweig

# Heterogeneities among credit risk parameter distributions: the modality defines the best estimation method

---

Marvin Zöllner

Department of Finance, University of Braunschweig - Institute of Technology



**Marvin Zöllner**



Abt-Jerusalem-Straße 7  
38106 Braunschweig  
Germany



+49 531 2894 /  
+49 15142878500



marvin.zoellner@tu-bs.de



## Education

### 2023 (planned)

Postdoctoral qualification, TU  
Braunschweig

### 2017 – 2022

Doctorate in finance,  
Braunschweig

### 2014– 2017

Fellowship for special  
achievements in thermodynamics

### 2013– 2014

Erasmus in Spain, Universidad  
Politécnica de Madrid

### 2012 – 2017

Studies in  
industrial/mechanical  
engineering, TU  
Braunschweig



## Studies with GCD Data

### 2023 (planned)

Portfolio optimization in credit risk management  
(together with M. Gürtler, E. Bondzio).

### 2023

Transparency in credit risk modeling:  
Improving linear regression  
(Working paper, together with M. Gürtler)

### 2022

Clustering meets machine learning: Increasing the  
accuracy of advanced tree-based methods in LGD  
estimation

(Working paper, SSRN, in submission process,  
together with M. Gürtler)

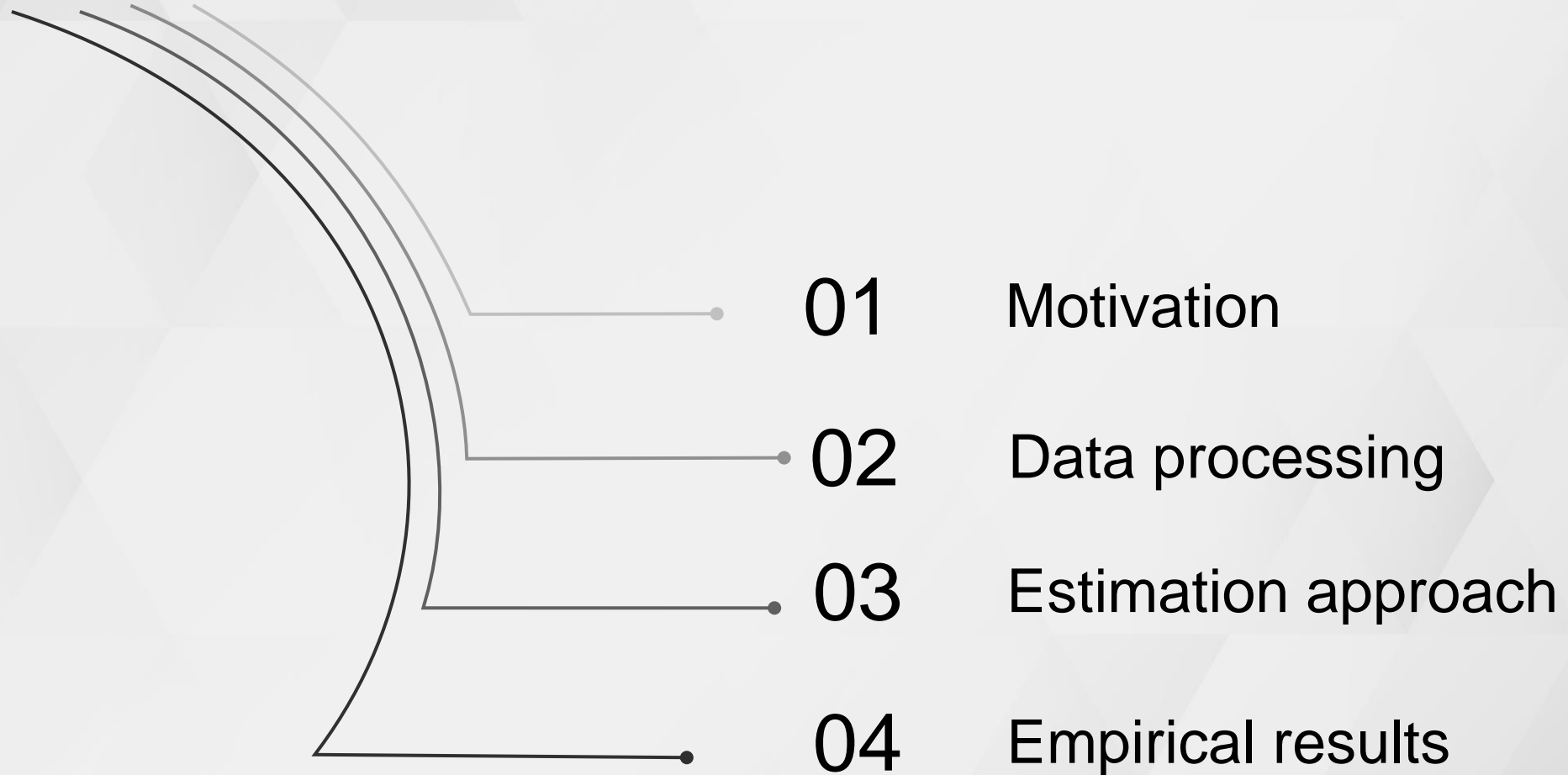
### 2022

Heterogeneities among credit risk parameter  
distributions: the modality defines the best estimation  
method

OR Spectrum, Vol. 44, coming soon  
(together with M. Gürtler).

# Agenda

---



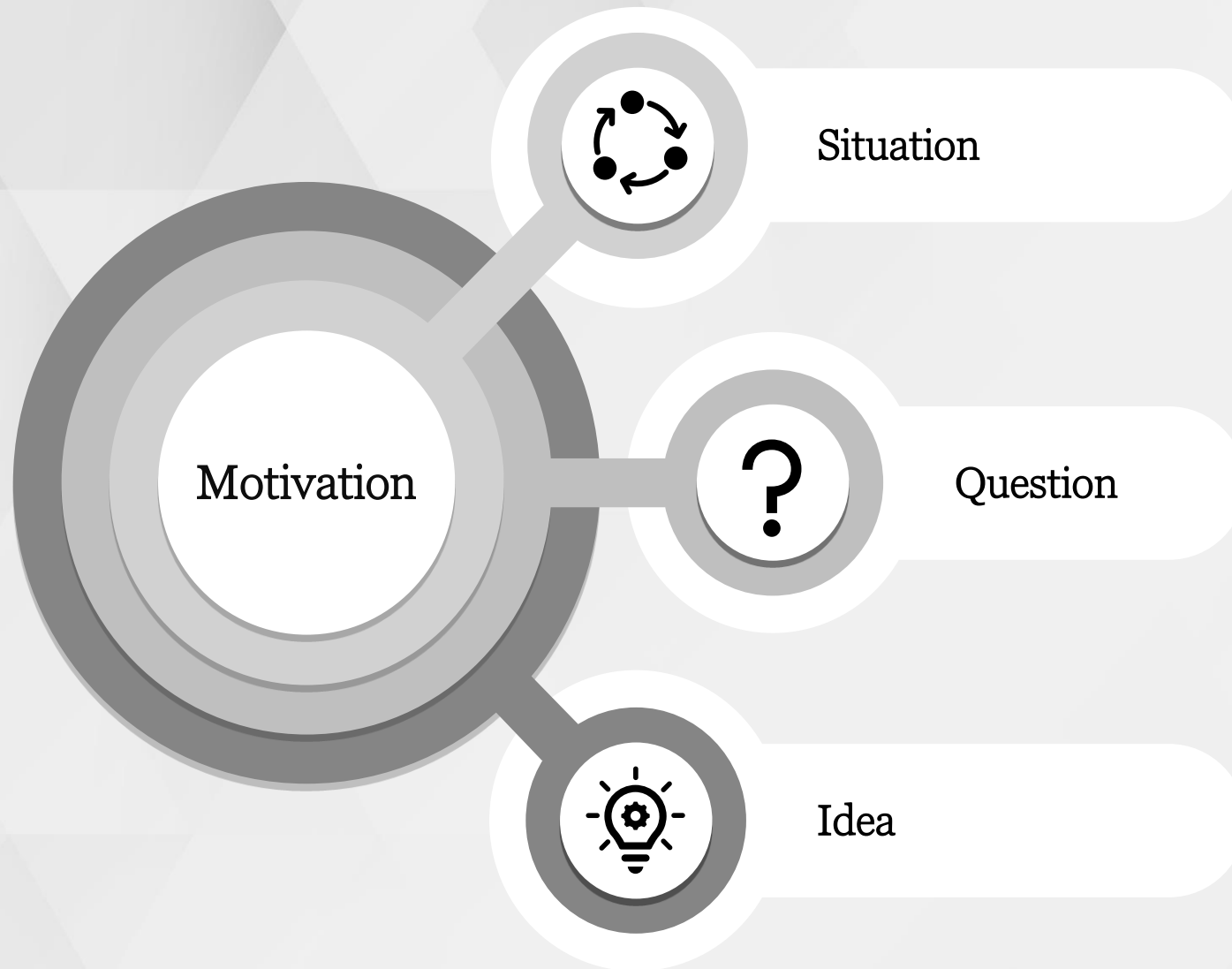
Authors	Journal	Data	Competitive methods	Measures	Best method
Hurlin et al. (2018)	European Journal of Operational Research	defaulted customers in <b>Brazil</b>	fractional response regression, regression tree, random forest, gradient boosting, artificial neural network, support vector regression	MSE, MAE	random forest
Kaposty et al.(2020)	International Journal of Forecasting	defaulted corporate leases in <b>Germany</b>	linear regression, tobit, regression tree, random forest, neural network,	MSE, MAE, $R^2$	random forest
Yao et al. (2017)	European Journal of Operational Research	defaulted credit cards in <b>United Kingdom</b>	linear regression, support vector regression, two-stage models, random forest	RMSE, MAE, $R^2$	support vector regression
Loterman et al. (2012)	International Journal of Forecasting	Defaulted personal, mortgage and corporate loans (location: <b>unknown</b> )	linear regression, beta-regression, regresstion tree, support vector regression, neural network, ...	RMSE, MAE, $R^2$	support vector regression / neural network
Tobback et al. (2014)	Journal of the Operational Research Society	Defaulted loans of home equity and corporates in <b>United States</b>	Linear regression, box-cox tranfomation, regression tree, support vector regression	Spearman's $\rho$ , RMSE, MAE, $R^2$	regression tree
Min et al. (2020)	Mathematics	Defaulted loans of SMEs in <b>United States</b>	regression models, decision tree, mixture regression models, neural network, ...	MSE	mixture regression models
Bellotti et. al (2021)	International Journal of Forecasting	Defaulted loans in <b>Europe</b>	Linear regression, regression splines, k-nearst neighbors, neural networks, vector regression, various tree-based methods, ...	RMSE, MAE	Gradient boosting

## Situation

- Mixed views on how well various LGD estimation methods perform.
- Studies look at different countries in which the loan portfolios are located.  
(see, e.g., [Bastos \(2010\)](#))
- National differences in bankruptcy law or the characteristics of borrowing companies.  
(see [Grunert and Weber \(2009\)](#))
- Differences in LGD characteristics across geographic regions.  
(see [Grippa et al. \(2005\)](#) and [Querci \(2005\)](#))

## Contribution

- LGD distributions seem to differ between regions and countries.
- Which characteristics of an LGD distribution are responsible for the different performance results?
- Identification of the distributional features relevant to the quality of LGD estimation methods.
- Determination of the methods that have the highest estimation accuracy for the relevant distribution types.



Variety of methods for LGD estimation (regression and machine learning) .

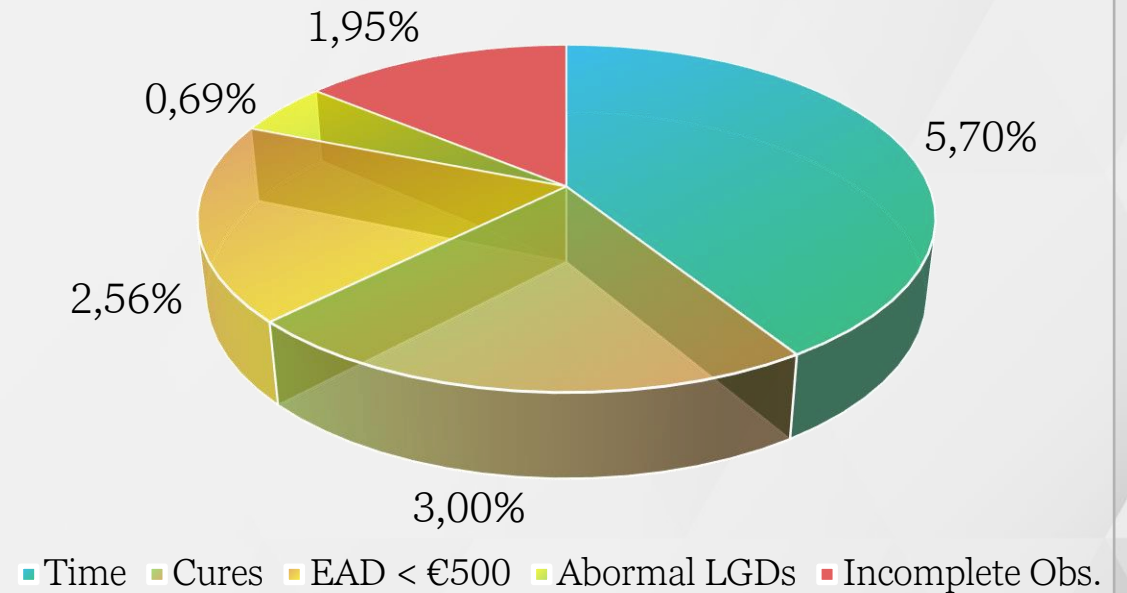
Which method should be used for LGD estimation? Which is best?

Comparison of estimation methods for a large class of LGD distributions.

# Data processing

- Dataset of resolved defaulted loans by SMEs from 16 European countries.
- Raw dataset comprises **38,166** defaulted loans.
- Filter rules:  
(Gürtler and Hibbeln (2013), European Bank Authority (2016), Krüger and Rösch (2017) and Betz et al. (2018))
  - 1) Data from 2000 to 2016
  - 2) No cures
  - 3) No loans with  $EAD < €500$
  - 4) No loans with abnormal high or low LGD (smaller -100%, higher than 200%)
  - 5) No incomplete observations
- Dataset of **32,851** loans remains.

$$LGD = 1 - \frac{\sum F^+ - \sum C^-}{EAD}$$



# Descriptive statistics

## Plausibility check ✓

- Guarantees or securities reduces LGDs.
- Non-senior and short-term loans lead to higher LGDs.
- Loans from the “finance, insurance, real estate” sector have the lowest LGDs.

**Table 1** Descriptive statistics

	Quantiles					Mean	Obs.
	0.05	0.25	0.50	0.75	0.95		
<i>LGD<sub>overall</sub></i>	-7.92	0.08	5.45	61.36	100.00	28.86	32851
<i>log(EAD)</i>	8.11	10.00	11.36	12.64	14.39	11.33	32851
Number of collaterals	0.00	0.00	1.00	1.00	3.00	1.18	32851
Number of guarantors	0.00	0.00	0.00	0.00	2.00	0.31	32851
<i>LGD conditional to guarantee availability:</i>							
No guarantee	-7.91	0.08	5.55	65.73	100.00	29.89	28774
Guarantee	-7.93	0.04	5.04	33.94	100.00	21.57	4077
<i>LGD conditional to collateral type:</i>							
No collateral	-2.62	1.88	24.09	98.22	100.00	43.37	9008
Real estate	-8.47	-0.55	2.12	26.37	99.45	18.13	7188
Other	-9.71	-0.09	4.61	51.09	100.00	25.64	16655
<i>LGD conditional to facility type:</i>							
Medium term	-3.99	0.25	4.69	48.95	100.00	26.26	19463
Short term	-15.54	-0.39	8.30	82.77	100.00	32.72	12658
Other	-6.80	0.08	4.43	79.98	100.00	31.35	730
<i>LGD conditional to seniority type:</i>							
Pari-passu	-8.47	0.01	5.64	60.03	100.00	28.63	27013
Super senior	-4.67	0.99	4.10	62.91	100.00	29.04	5301
Non-senior	-9.81	0.73	20.65	84.76	100.00	38.40	537
<i>LGD conditional to industry type:</i>							
Finance, insurance, real estate	-8.74	-0.48	1.95	45.30	100.00	23.18	4726
Agriculture, forestry, fishing, hunting	-6.17	-0.66	2.43	66.13	100.61	27.20	1486
Mining	-5.52	0.65	1.13	53.88	100.00	26.82	120
Construction	-10.42	-0.02	3.54	62.49	100.00	27.74	3720
Manufacturing	-10.29	0.01	5.37	64.12	100.00	29.35	4707
Transp., commu., elec., gas, sani. serv.	-6.76	0.61	2.47	51.53	100.00	25.08	2587
Wholesale and retail trade	-7.82	0.08	6.28	74.09	100.00	32.42	5483
Services	-6.85	0.82	14.41	76.85	100.00	34.12	5877
Other	-5.18	0.84	10.27	45.26	100.00	26.64	4145

Note. This table presents the means and quantiles of empirical LGDs (in %) for different loan categories

# Estimation methods

20 different methods, categorized as either traditional or advanced methods.

- Linear regression as reference method
- Different variable selection techniques
- Tree-based methods
- Boosting methods
- Advanced methods (e.g., neural network, SVR)

**Table 2** Competing methods

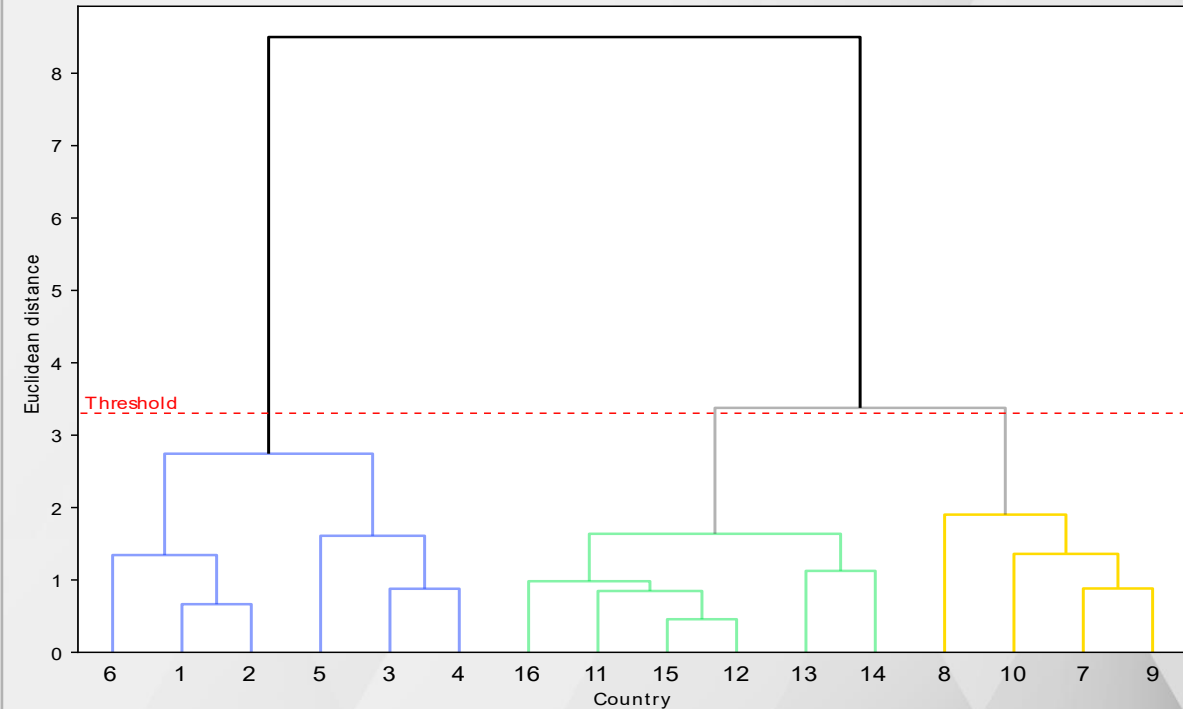
Method	Exemplary literature
Ordinary Least Squares ( <i>OLS</i> )	Qi and Zhao (2011), Krüger and Rösch (2017)
<i>Variable selection methods:</i>	
(1) OLS with Backward Elimination ( <i>bOLS</i> )	Hartmann-Wendels et al. (2014), Ye and Bellotti (2019)
(2) Least Angle Regression (LAR)	
<i>Penalized Regressions:</i>	
(1) Ridge Regression ( <i>RR</i> )	Loterman et al. (2012)
(2) Lasso Regression ( <i>LR</i> )	
(3) Elastic Regression ( <i>ER</i> )	
Fractional Logit Regression ( <i>FLR</i> )	Dermine and de Carvalho (2006), Chava et al. (2011)
Regression Tree ( <i>RT</i> )	Matuszyk et al. (2010), Hurlin et al. (2018)
Conditional Inference Tree ( <i>CIT</i> )	Hothorn et al. (2006), Bellotti et al. (2021)
Random Forest ( <i>RF</i> )	Miller and Töws (2018), Hurlin et al. (2018)
<i>Boosting Methods:</i>	
(1) Adaptive Boosting ( <i>ADA</i> ),	Tanoue and Yamashita (2019)
(2) Gradient Boosting ( <i>GB</i> )	
Cubist Regression Model ( <i>CUB</i> )	Kuhn and Quinlan (2018), Bellotti et al. (2021)
Artificial Neural Network ( <i>ANN</i> )	Qi and Zhao (2011), Hurlin et al. (2018)
Support Vector Regression ( <i>SVR</i> )	Yao et al. (2015, 2017), Nazemi et al. (2017)
Relevance Vector Regression ( <i>RVR</i> )	Karatzoglou et al. (2004), Bellotti et al. (2021)
Gaussian Process Regression ( <i>GAPR</i> )	Bellotti et al. (2021)
K-nearest Neighbors ( <i>KNN</i> )	Yang and Tkachenko (2012), Hartmann-Wendels et al. (2014)
Multivariate Adaptive Regression Splines ( <i>MARS</i> )	Loterman et al. (2012), Bellotti et al. (2021)
Finite Mixture Model ( <i>FMM</i> )	Krüger and Rösch (2017), Min et al. (2020)

# Clustering and LGD distribution analysis

## Aims

- 1) Identification of country-specific types of LGD distributions from 16 European countries.
- 2) Application of cluster analysis to identify relevant types of distributions.
- 3) Building of country-specific subsamples for the respective distribution clusters.
- 4) Comparison of the LGD methods to identify the method with the highest accuracy for each subsample.

For each country, we aggregate the LGDs of all defaulted loans based on the LGD quantiles in a range from 1% to 100% with a stepwise increase of 1%.



### Cluster 1

Czech Republic, Denmark, Lithuania, Norway,  
Poland, Romania

→ Symmetric bimodal distribution.

### Cluster 2

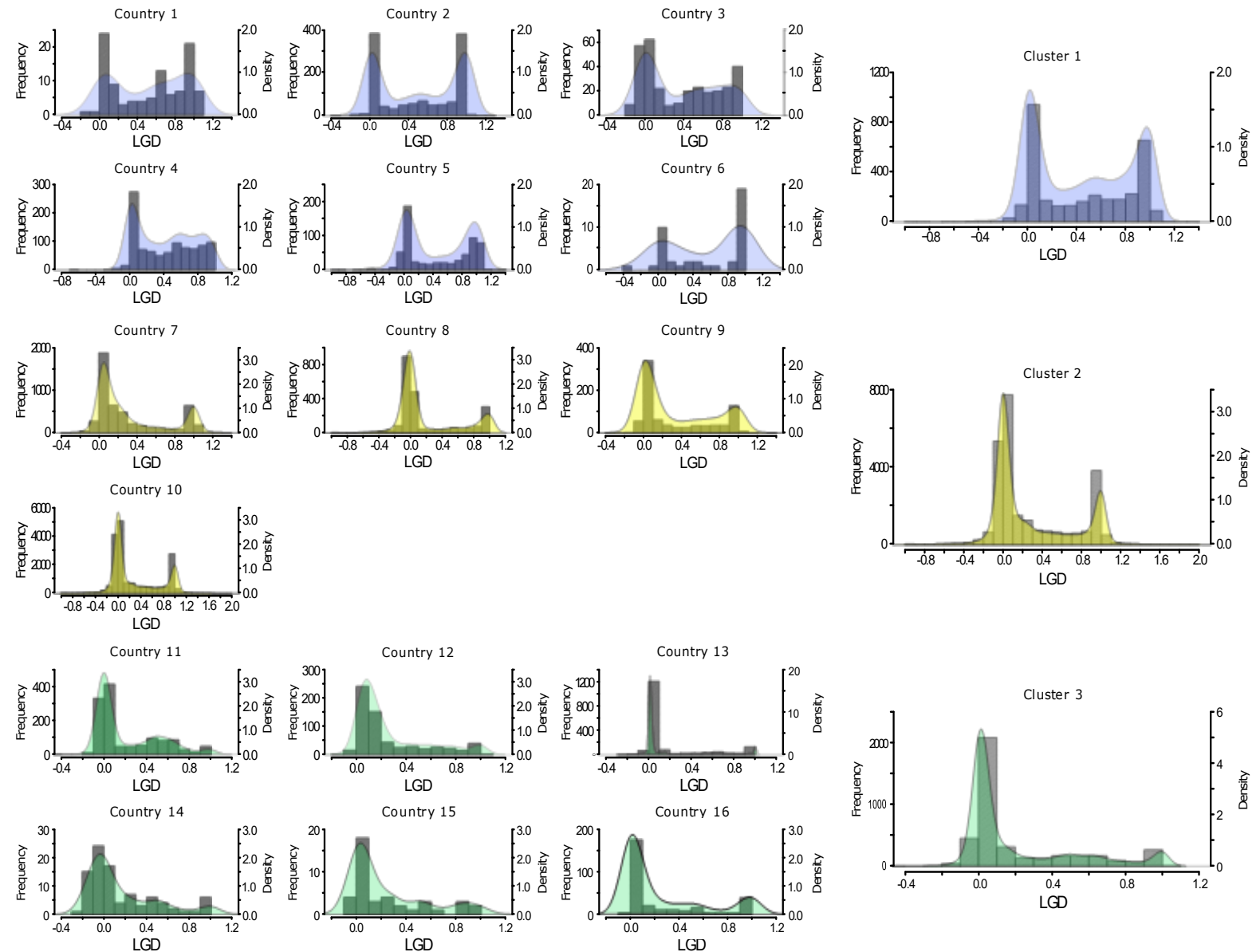
Germany, Ireland, Sweden, United Kingdom

→ Asymmetric (positively skewed) bimodal distribution.

### Cluster 3

Estonia, France, Latvia, Netherlands, Switzerland,  
Finland

→ (Positively skewed) unimodal distribution.



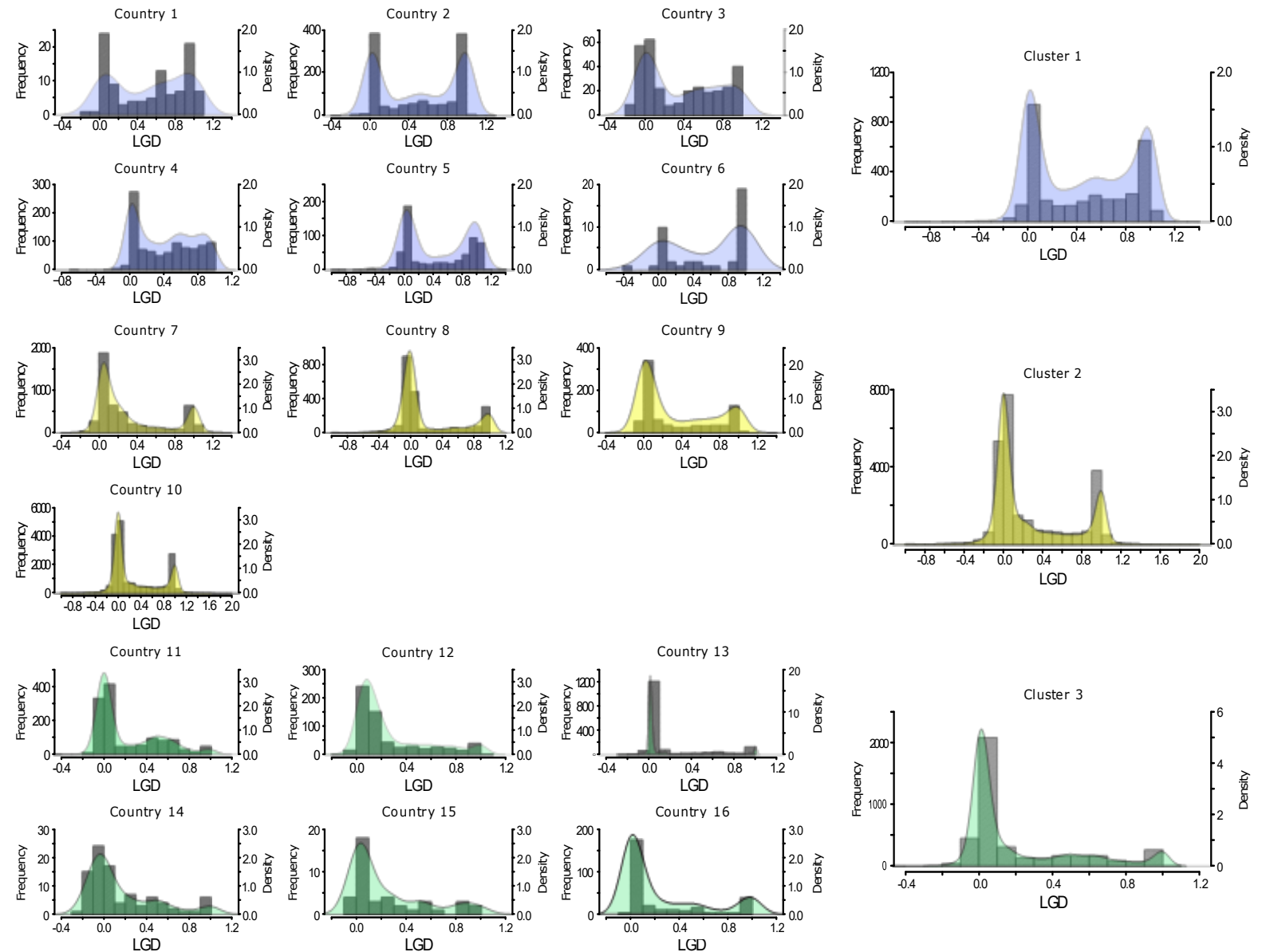
### Paired t-tests

	Cluster 1	Cluster 2	Cluster 3
Cluster 1	-		
Cluster 2	-12.1608***	-	
Cluster 3	22.5953***	30.9241***	-

### Paired Mann-Whitney U tests

	Cluster 1	Cluster 2	Cluster 3
Cluster 1	-		
Cluster 2	5.1E+07***	-	
Cluster 3	3.0E+07***	4.1E+06***	-

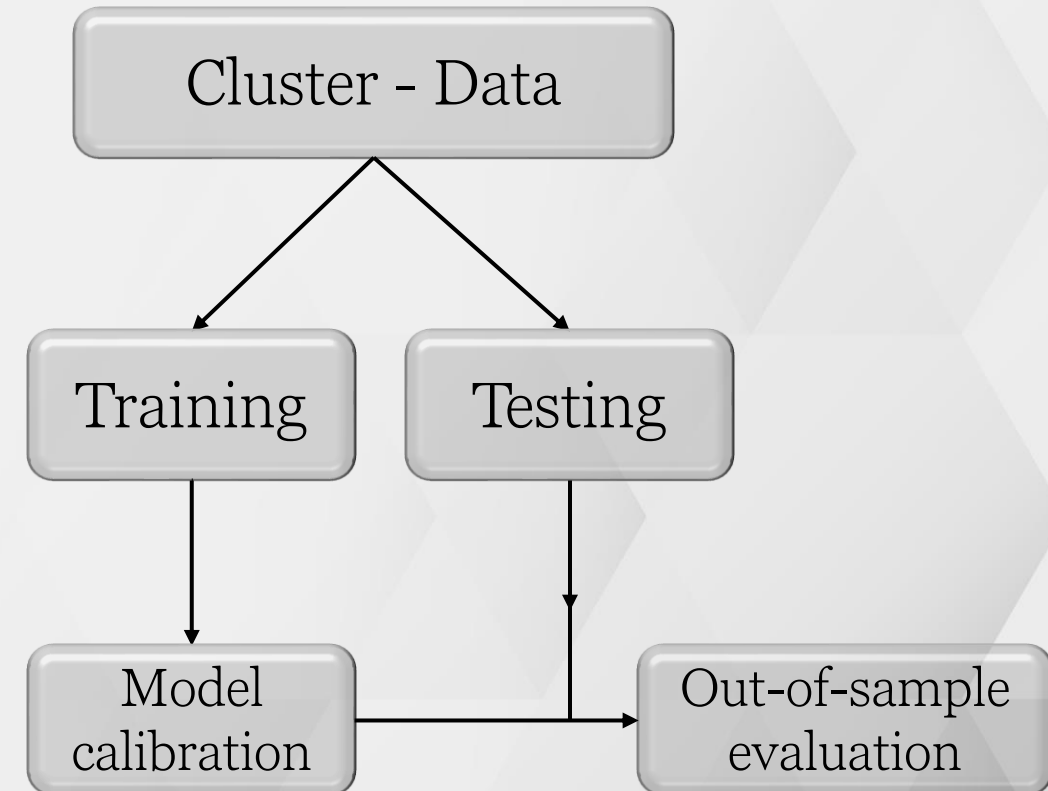
Note. Values are paired t statistics and U statistics, respectively. †, \*, \*\*, and \*\*\* imply significance at the 10%, 5%, 1%, and 0.1% levels.



# Model comparison procedure

- Procedure is performed for each of the three cluster subsamples.
- Different split ratios.
  - (60/40), (70/30), (80/20), (90/10)
- Different out-of-sample/time criteria.
  - MSE, MAE,  $R^2$
- Mean of each criteria calculated over all sample splits denotes the final estimation accuracy.
- Intensive Hyperparametertuning by five-fold-cross-validation.

1	2	3	4	5
Train	Train	Test	Train	Train



# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	<u>0.1241</u>	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	<u>0.1276</u>	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	<u>0.0455</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*)

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	<u>0.1241</u>	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	<u>0.1276</u>	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	<u>0.0455</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad f.t. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*)

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	<u>0.1241</u>	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	<u>0.1276</u>	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	<u>0.0807</u>	<u>0.0813</u>	<u>0.0840</u>	<u>0.0808</u>	<u>0.0807</u>	<u>0.0807</u>	<u>0.0826</u>	<u>0.0677</u>	<u>0.0650</u>	<u>0.0561</u>	<u>0.0758</u>	<u>0.0572</u>	<u>0.0600</u>	<u>0.0788</u>	<u>0.0602</u>	<u>0.0647</u>	<u>0.0633</u>	<u>0.0636</u>	<u>0.0683</u>	<u>0.0455</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad f.t. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	<i>OLS</i>	<i>bOLS</i>	<i>LAR</i>	<i>RR</i>	<i>LR</i>	<i>ER</i>	<i>FLR</i>	<i>RT</i>	<i>CIT</i>	<i>RF</i>	<i>ADA</i>	<i>GB</i>	<i>CUB</i>	<i>ANN</i>	<i>SVR</i>	<i>RVR</i>	<i>GAPR</i>	<i>KNN</i>	<i>MARS</i>	<i>FMM</i>
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	<u>0.1241</u>	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	<u>0.1276</u>	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	<u>0.0807</u>	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	<u>0.0758</u>	0.0572	0.0600	<u>0.0788</u>	0.0602	0.0647	0.0633	0.0636	0.0683	<u>0.0455</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MSE)

**Table 4** European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1338	0.1344	0.1336	0.1349	0.1330	0.1316	0.1330	0.1350	0.1379	0.1268	0.1361	0.1321	0.1345	0.1374	0.1354	0.1372	0.1357	0.1463	0.1360	0.1350
70/30	0.1359	0.1355	0.1362	0.1332	0.1347	0.1348	0.1343	0.1353	0.1320	0.1235	0.1347	0.1320	0.1314	0.1346	0.1321	0.1363	0.1323	0.1442	0.1337	0.1327
80/20	0.1316	0.1332	0.1317	0.1329	0.1329	0.1346	0.1341	0.1333	0.1294	0.1238	0.1315	0.1308	0.1322	0.1314	0.1326	0.1311	0.1317	0.1419	0.1342	0.1320
90/10	0.1311	0.1299	0.1309	0.1324	0.1326	0.1326	0.1324	0.1302	0.1309	0.1221	0.1303	0.1270	0.1317	0.1311	0.1285	0.1282	0.1309	0.1399	0.1282	0.1314
Mean	0.1331	0.1332	0.1331	0.1334	0.1333	0.1334	0.1335	0.1335	0.1326	0.1241	0.1332	0.1305	0.1324	0.1336	0.1322	0.1332	0.1327	0.1431	0.1330	0.1328
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1388	0.1383	0.1388	0.1382	0.1383	0.1380	0.1369	0.1357	0.1352	0.1330	0.1474	0.1287	0.1351	0.1356	0.1348	0.1358	0.1340	0.1358	0.1351	0.1354
70/30	0.1376	0.1376	0.1376	0.1376	0.1376	0.1378	0.1357	0.1344	0.1340	0.1324	0.1462	0.1282	0.1341	0.1346	0.1343	0.1345	0.1335	0.1345	0.1344	0.1349
80/20	0.1366	0.1366	0.1381	0.1373	0.1376	0.1378	0.1359	0.1339	0.1334	0.1325	0.1443	0.1273	0.1335	0.1340	0.1332	0.1343	0.1331	0.1344	0.1354	0.1334
90/10	0.1353	0.1361	0.1364	0.1360	0.1361	0.1366	0.1360	0.1331	0.1317	0.1296	0.1438	0.1260	0.1305	0.1335	0.1303	0.1337	0.1309	0.1335	0.1318	0.1316
Mean	0.1371	0.1371	0.1377	0.1373	0.1374	0.1375	0.1361	0.1343	0.1336	0.1319	0.1454	0.1276	0.1333	0.1344	0.1331	0.1346	0.1329	0.1345	0.1342	0.1338
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0828	0.0815	0.0857	0.0810	0.0817	0.0819	0.0847	0.0682	0.0658	0.0603	0.0777	0.0607	0.0642	0.0805	0.0642	0.0660	0.0644	0.0646	0.0689	0.0496
70/30	0.0813	0.0812	0.0825	0.0809	0.0807	0.0799	0.0825	0.0675	0.0649	0.0551	0.0773	0.0576	0.0599	0.0791	0.0603	0.0636	0.0627	0.0627	0.0689	0.0412
80/20	0.0795	0.0817	0.0837	0.0804	0.0801	0.0802	0.0829	0.0677	0.0646	0.0549	0.0761	0.0568	0.0591	0.0787	0.0598	0.0656	0.0636	0.0630	0.0682	0.0455
90/10	0.0790	0.0808	0.0843	0.0809	0.0803	0.0807	0.0805	0.0674	0.0647	0.0540	0.0720	0.0536	0.0570	0.0768	0.0565	0.0636	0.0624	0.0642	0.0674	0.0456
Mean	0.0807	0.0813	0.0840	0.0808	0.0807	0.0807	0.0826	0.0677	0.0650	0.0561	0.0758	0.0572	0.0600	0.0788	0.0602	0.0647	0.0633	0.0636	0.0683	0.0455

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit. The methods are abbreviated as follows: Ordinary Least Squares (*OLS*); *OLS* with Backward Elimination (*bOLS*); Least Angle Regression (*LAR*); Ridge Regression (*RR*); Lasso Regression (*LR*); Elastic Regression (*ER*); Fractional Logit Regression (*FLR*); Regression Tree (*RT*); Conditional Inference Tree (*CIT*); Random Forest (*RF*); Adaptive Boosting (*ADA*); Gradient boosting (*GB*); Cubist Regression Model (*CUB*); Artificial Neural Network (*ANN*); Support Vector Regression (*SVR*); Relevance Vector Regression (*RVR*); Gaussian Processes (*GAPR*); K-nearest Neighbors (*KNN*); Multivariate Adaptive Regression Splines (*MARS*); Finite Mixture Model (*FMM*).

# Out-of-sample/time results (MAE)

**Table 5** European credit portfolios: Out-of-sample estimation accuracies (MAE)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.3225	0.3234	0.3244	0.3217	0.3278	0.3244	0.3107	0.3178	0.3296	0.3128	0.3354	0.3134	0.3099	0.2969	0.3279	0.3209	0.3221	0.3391	0.3271	0.3206
70/30	0.3250	0.3253	0.3254	0.3237	0.3296	0.3284	0.3311	0.3247	0.3147	0.3075	0.3339	0.3133	0.3057	0.3140	0.3236	0.3106	0.3182	0.3354	0.3210	0.3419
80/20	0.3188	0.3221	0.3201	0.3208	0.3276	0.3286	0.3253	0.3156	0.3097	0.3077	0.3327	0.3115	0.3065	0.3090	0.3242	0.3052	0.3178	0.3352	0.3199	0.3205
90/10	0.3176	0.3169	0.3172	0.3196	0.3261	0.3245	0.3298	0.3115	0.3099	0.2989	0.3323	0.3129	0.3053	0.3229	0.3170	0.3014	0.3186	0.3278	0.3118	0.3197
Mean	0.3210	0.3219	0.3218	0.3214	0.3278	0.3265	0.3242	0.3174	0.3160	0.3067	0.3336	0.3128	0.3068	0.3107	0.3232	0.3095	0.3192	0.3343	0.3200	0.3257
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.2963	0.2953	0.2958	0.2953	0.2975	0.2959	0.3017	0.2844	0.2843	0.2887	0.3430	0.2731	0.2736	0.2940	0.2780	0.2826	0.2861	0.2807	0.2867	0.3470
70/30	0.2950	0.2950	0.2950	0.2950	0.2970	0.2966	0.2929	0.2817	0.2815	0.2884	0.3434	0.2745	0.2732	0.2659	0.2779	0.2811	0.2869	0.2795	0.2868	0.3467
80/20	0.2924	0.2924	0.2951	0.2940	0.2971	0.2962	0.2954	0.2806	0.2789	0.2881	0.3388	0.2697	0.2705	0.2604	0.2751	0.2796	0.2859	0.2776	0.2869	0.3450
90/10	0.2923	0.2939	0.2945	0.2939	0.2954	0.2956	0.2570	0.2785	0.2764	0.2853	0.3394	0.2676	0.2678	0.2847	0.2712	0.2794	0.2829	0.2780	0.2845	0.3411
Mean	0.2940	0.2942	0.2951	0.2946	0.2968	0.2961	0.2867	0.2813	0.2803	0.2876	0.3411	0.2712	0.2713	0.2763	0.2756	0.2807	0.2854	0.2790	0.2862	0.3450
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.2130	0.2077	0.2238	0.2067	0.2091	0.2108	0.2217	0.1641	0.1679	0.1518	0.2116	0.1599	0.1603	0.2228	0.2143	0.1691	0.1679	0.1633	0.1704	0.1499
70/30	0.2123	0.2112	0.2162	0.2111	0.2091	0.2067	0.2162	0.1621	0.1661	0.1498	0.1961	0.1569	0.1521	0.1944	0.2097	0.1648	0.1669	0.1498	0.1687	0.1493
80/20	0.2047	0.2121	0.2202	0.2079	0.2045	0.2051	0.2163	0.1636	0.1675	0.1440	0.2288	0.1527	0.1528	0.1940	0.2095	0.1699	0.1706	0.1563	0.1737	0.1470
90/10	0.1997	0.2046	0.2183	0.2068	0.2022	0.2036	0.2408	0.1666	0.1763	0.1496	0.2318	0.1574	0.1557	0.2305	0.2071	0.1684	0.1684	0.1787	0.1725	0.1460
Mean	0.2074	0.2089	0.2196	0.2081	0.2062	0.2065	0.2238	0.1641	0.1694	0.1488	0.2171	0.1568	0.1552	0.2104	0.2101	0.1680	0.1685	0.1620	0.1713	0.1480

Note. This table reports the MAE of the out-of-sample estimation for the considered LGD methods. High values for MAE imply a bad fit

# Out-of-sample/time results ( $R^2$ )

Table OA.4: European credit portfolios: Out-of-sample estimation accuracies ( $R^2$ ).

Note. This table reports the  $R^2$  of the out-of-sample estimation for the considered LGD methods. Low values for  $R^2$  imply a bad fit.

Cluster 1: (nearly) symmetric bimodal LGD distribution

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
60/40	0.1733	0.1828	0.1715	0.1794	0.1807	0.1799	0.1828	0.1788	0.1613	0.2286	0.1719	0.1962	0.1817	0.1645	0.1765	0.1652	0.1746	0.1102	0.1729	0.1786
70/30	0.1769	0.1664	0.1778	0.1805	0.1813	0.1903	0.1818	0.1672	0.1878	0.2402	0.1709	0.1879	0.1916	0.1720	0.1872	0.1614	0.1860	0.1127	0.1771	0.1835
80/20	0.1960	0.1865	0.1953	0.1878	0.1882	0.1776	0.1806	0.1854	0.2092	0.2438	0.1963	0.2012	0.1925	0.1970	0.1897	0.1992	0.1954	0.1331	0.1798	0.1939
90/10	0.2093	0.2169	0.2108	0.2016	0.2008	0.2004	0.2019	0.2147	0.2105	0.2636	0.2145	0.2340	0.2059	0.2093	0.2251	0.2269	0.2107	0.1565	0.2269	0.2078
Mean	0.1889	0.1881	0.1889	0.1873	0.1878	0.1870	0.1868	0.1865	0.1922	0.2440	0.1884	0.2048	0.1930	0.1857	0.1946	0.1882	0.1916	0.1281	0.1892	0.1909

Cluster 2: asymmetric (positively skewed) bimodal LGD distribution

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
60/40	0.1934	0.1964	0.1935	0.1965	0.1964	0.1980	0.2043	0.2115	0.2140	0.2268	0.1435	0.2519	0.2147	0.2121	0.2168	0.2107	0.2214	0.2110	0.2148	0.2133
70/30	0.1991	0.1992	0.1991	0.1991	0.1993	0.1982	0.2103	0.2176	0.2202	0.2293	0.1490	0.2537	0.2198	0.2165	0.2187	0.2170	0.2231	0.2174	0.2181	0.2148
80/20	0.2036	0.2032	0.1945	0.1991	0.1976	0.1966	0.2074	0.2193	0.2221	0.2274	0.1585	0.2576	0.2214	0.2185	0.2235	0.2170	0.2238	0.2161	0.2107	0.2221
90/10	0.2167	0.2119	0.2100	0.2122	0.2120	0.2088	0.2121	0.2293	0.2374	0.2495	0.1670	0.2701	0.2440	0.2270	0.2452	0.2257	0.2420	0.2267	0.2367	0.2379
Mean	0.2032	0.2027	0.1993	0.2017	0.2013	0.2004	0.2086	0.2194	0.2234	0.2332	0.1545	0.2583	0.2249	0.2185	0.2261	0.2176	0.2276	0.2178	0.2201	0.2220

Cluster 3: (positively skewed) unimodal LGD distribution

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
60/40	0.1357	0.1497	0.1052	0.1545	0.1474	0.1450	0.1163	0.2877	0.3134	0.3708	0.1892	0.3670	0.3301	0.1597	0.3303	0.3115	0.3283	0.3258	0.2812	0.4825
70/30	0.1463	0.1479	0.1342	0.1503	0.1522	0.1612	0.1339	0.2909	0.3189	0.4210	0.1887	0.3952	0.3711	0.1695	0.3664	0.3326	0.3421	0.3422	0.2763	0.5674
80/20	0.1793	0.1563	0.1356	0.1702	0.1731	0.1727	0.1445	0.3009	0.3335	0.4329	0.2148	0.4134	0.3899	0.1878	0.3832	0.3225	0.3434	0.3495	0.2965	0.5300
90/10	0.2033	0.1845	0.1500	0.1837	0.1901	0.1855	0.1882	0.3204	0.3472	0.4553	0.2735	0.4590	0.4252	0.2256	0.4296	0.3579	0.3706	0.3520	0.3199	0.5404
Mean	0.1661	0.1596	0.1313	0.1647	0.1657	0.1661	0.1457	0.3000	0.3282	0.4200	0.2165	0.4086	0.3791	0.1856	0.3773	0.3311	0.3461	0.3424	0.2935	0.5301

# Significance test (MSE)

- Paired t-test to exclude the possibility that superiority may have occurred by chance.
- Cluster 1-3:

the differences in mean MSE between best and next best methods **are negative at the 5% significance level.**

Note. Values are paired t statistics where a negative value means the mean MSE (that is, the MSEs of all splits are considered) for the method on the vertical axis is worse than that for the method on the horizontal axis, and vice versa. †, \*, \*\*, and \*\*\* imply significance at the 10%, 5%, 1%, and 0.1% levels.

Cluster 1: (nearly) symmetric bimodal LGD distribution

	RF	GB	SVR	CUB	CIT
RF	-				
GB	-2.0898*	-			
SVR	-2.4173*	-0.5926	-		
CUB	-2.4362*	-0.4377	-0.1006	-	
CIT	-2.8967**	-0.7428	-0.1859	-0.2637	-

Cluster 2: asymmetric (positively skewed) bimodal LGD distribution

	GB	RF	GAPR	SVR	CUB
GB	-				
RF	-2.4225*	-			
GAPR	-2.9307**	-0.5263	-		
SVR	-3.1412**	-0.7891	-0.2727	-	
CUB	-3.1659**	-0.8689	-0.3644	-0.0963	-

Cluster 3: (positively skewed) unimodal LGD distribution

	FMM	RF	GB	CUB	SVR
FMM	-				
RF	-4.3298***	-			
GB	-4.9064***	-0.4211	-		
CUB	-5.8791***	-1.3845	-0.9880	-	
SVR	-7.4937***	-1.7184	-1.2601	-0.0848	-

# Robustness checks

Four modifications of the estimation approach:

- 1) extend the methods by including additional firm-specific explanatory variables (e.g., total assets).
- 2) change the clustering procedure by clustering the LGD distributions based on a loan-specific variable.
- 3) apply a logarithmic transformation to the positively skewed unimodally distributed LGDs.
- 4) use non-European credit portfolios that are characterized by the same three types of LGD distributions.



The best methods remain the same for the three distribution types.

# Conclusion

- The literature reveals mixed results on how well LGD estimation methods perform.
  - We compare various LGD estimation methods for a large class of LGD distributions.
  - The cluster analysis leads to three types of distributions, which differ in their modality.
  - The estimation accuracies of 20 different methods are tested based on their out-of-sample performance.
- 
- Random forest implies the highest estimation accuracy for (nearly) symmetric bimodal LGD distribution.
  - Gradient boosting shows best performance for asymmetric (positively skewed) bimodal LGD distribution.
  - Finite mixture model is best for (positively skewed) unimodal distribution.



Technische  
Universität  
Braunschweig

# Thank you for your attention

---

Marvin Zöllner

Department of Finance, University of Braunschweig - Institute of Technology

# References

---

- Bastos JA (2010) Forecasting bank loans loss-given-default. J Bank Financ 34(10):2510–2517. <https://doi.org/10.1016/j.jbankfin.2010.04.011>.
- Bellotti A, Brigo D, Gambetti P et al. (2021) Forecasting recovery rates on non-performing loans with machine learning. Int J Forecast 37(1):428–444. <https://doi.org/10.1016/j.ijforecast.2020.06.009>.
- Betz J, Kellner R, Rösch D (2018) Systematic effects among loss given defaults and their implications on downturn estimation. Eur J Oper Res 271(3):1113–1144. <https://doi.org/10.1016/j.ejor.2018.05.059>.
- Chava S, Stefanescu C, Turnbull S (2011) Modeling the loss distribution. Manag Sci 57(7):1267–1287.
- Dermine J, de Carvalho CN (2006) Bank loan losses-given-default: a case study. J Bank Financ 30(4):1219–1243. <https://doi.org/10.1016/j.jbankfin.2005.05>.
- European Banking Authority (2016) Guidelines on PD estimation. LGD estimation and the treatment of defaulted exposures, Consultation Paper.
- Grippa P, Iannotti S, Leandri F (2005) Recovery rates in the banking industry: stylised facts emerging from the Italian experience. Next Chall Credit Risk Manag Recovery Risk 121141:121–141.

# References

---

- Grunert J, Weber M (2009) Recovery rates of commercial lending: empirical evidence for german companies. J Bank Financ 33(3):505–513. <https://doi.org/10.1016/j.jbankfin.2008.09.002>.
- Gürtler M, Hibbeln M (2013) Improvements in loss given default forecasts for bank loans. J Bank Financ 37(7):2354–2366. <https://doi.org/10.1016/j.jbankfin.2013.01.031>.
- Hartmann-Wendels T, Miller P, Töws E (2014) Loss given default for leasing: parametric and nonparametric estimations. J Bank Financ 40:364–375. <https://doi.org/10.1016/j.jbankfin.2013.12.006>.
- Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. Journal of computational and graphical statistics 15(3):651–674. <https://doi.org/10.1198/106186006X133933>.
- Hurlin C, Leymarie J, Patin A (2018) Loss functions for loss given default model comparison. Eur J Oper Res 268(1):348–360. <https://doi.org/10.1016/j.ejor.2018.01.020>.
- Kaposty F, Kriebel J, Löderbusch M (2020) Predicting loss given default in leasing: a closer look at models and variable selection. Int J Forecast 36(2):248–266. <https://doi.org/10.1016/j.ijforecast.2019.05.009>.
- Karatzoglou A, Smola A, Hornik K (2004) kernlab – an s4 package for kernel methods in R. J Stat Softw. <https://doi.org/10.18637/jss>.

# References

---

- Krüger S, Rösch D (2017) Downturn LGD modeling using quantile regression. J Bank Financ 79:42–56. [https:// doi. org/ 10. 1016/j. jbank fin. 2017. 03. 001](https://doi.org/10.1016/j.jbankfin.2017.03.001) v011. i09.
- Kuhn M, Quinlan R (2018) Rule- and instance-based regression modeling. R Package Vers 2:1–14.
- Loterman G, Brown I, Martens D et al. (2012) Benchmarking regression algorithms for loss given default modeling. Int J Forecast 28(1):161–170. [https:// doi. org/ 10. 1016/j. ijfor ecast. 2011. 01](https://doi.org/10.1016/j.ijforecast.2011.01).
- Matuszyk A, Mues C, Thomas LC (2010) Modelling LGD for unsecured personal loans: decision tree approach. J Oper Res Soc 61(3):393–398. [https:// doi. org/ 10. 1057/ jors. 2009. 67](https://doi.org/10.1057/jors.2009.67).
- Miller P, Töws E (2018) Loss given default adjusted workout processes for leases. J Bank Financ 91:189–201. [https:// doi. org/ 10. 1016/j. jbank fin. 2017. 01. 020](https://doi.org/10.1016/j.jbankfin.2017.01.020).
- Min A, Scherer M, Schischke A et al. (2020) Modeling recovery rates of small- and medium-sized entities in the US. Mathematics 8(11):1856. [https:// doi. org/ 10. 3390/ math8 111856](https://doi.org/10.3390/math8111856).
- Qi M, Zhao X (2011) Comparison of modeling methods for loss given default. J Bank Financ 35(11):2842–2855. [https:// doi. org/ 10. 1016/j. jbank fin. 2011. 03. 011](https://doi.org/10.1016/j.jbankfin.2011.03.011).

# References

---

- Querci F (2005) Loss given default on a medium-sized Italian bank's loans: an empirical exercise. European financial management association.
- Tanoue Y, Yamashita S (2019) Loss given default estimation: a two-stage model with classification tree-based boosting and support vector logistic regression. J Risk 21(4):19–37. <https://doi.org/10.21314/JOR.2019.405>.
- Tobback E, Martens D, van Gestel T et al. (2014) Forecasting loss given default models: impact of account characteristics and the macroeconomic state. J Oper Res Soc 65(3):376–392. <https://doi.org/10.1057/jors.2013.158>.
- Yao X, Crook J, Andreeva G (2015) Support vector regression for loss given default modelling. Eur J Oper Res 240(2):528–538. <https://doi.org/10.1016/j.ejor.2014.06.043>.
- Yao X, Crook J, Andreeva G (2017) Enhancing two-stage modelling methodology for loss given default with support vector machines. Eur J Oper Res 263(2):679–689. <https://doi.org/10.1016/j.ejor.2017.05.017>.
- Ye H, Bellotti A (2019) Modelling recovery rates for non-performing loans. Risks 7(1):1–17. <https://doi.org/10.3390/risks7010019>.



Technische  
Universität  
Braunschweig

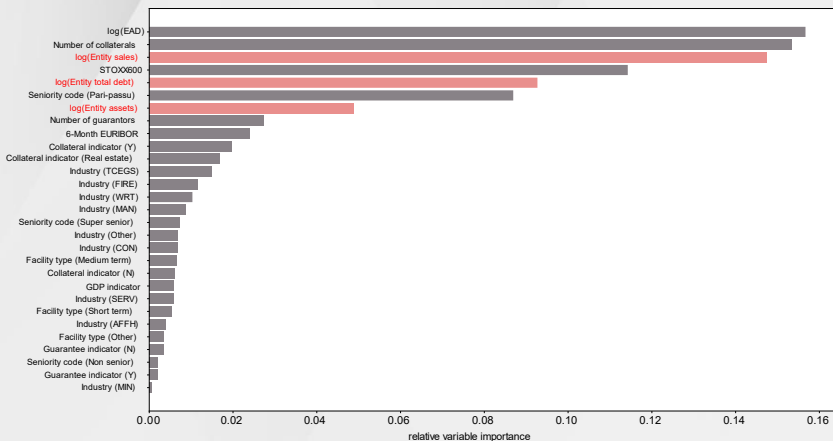
# Backup

---

Marvin Zöllner

Department of Finance, University of Braunschweig - Institute of Technology

# Robustness check: Inclusion firm-specific variables

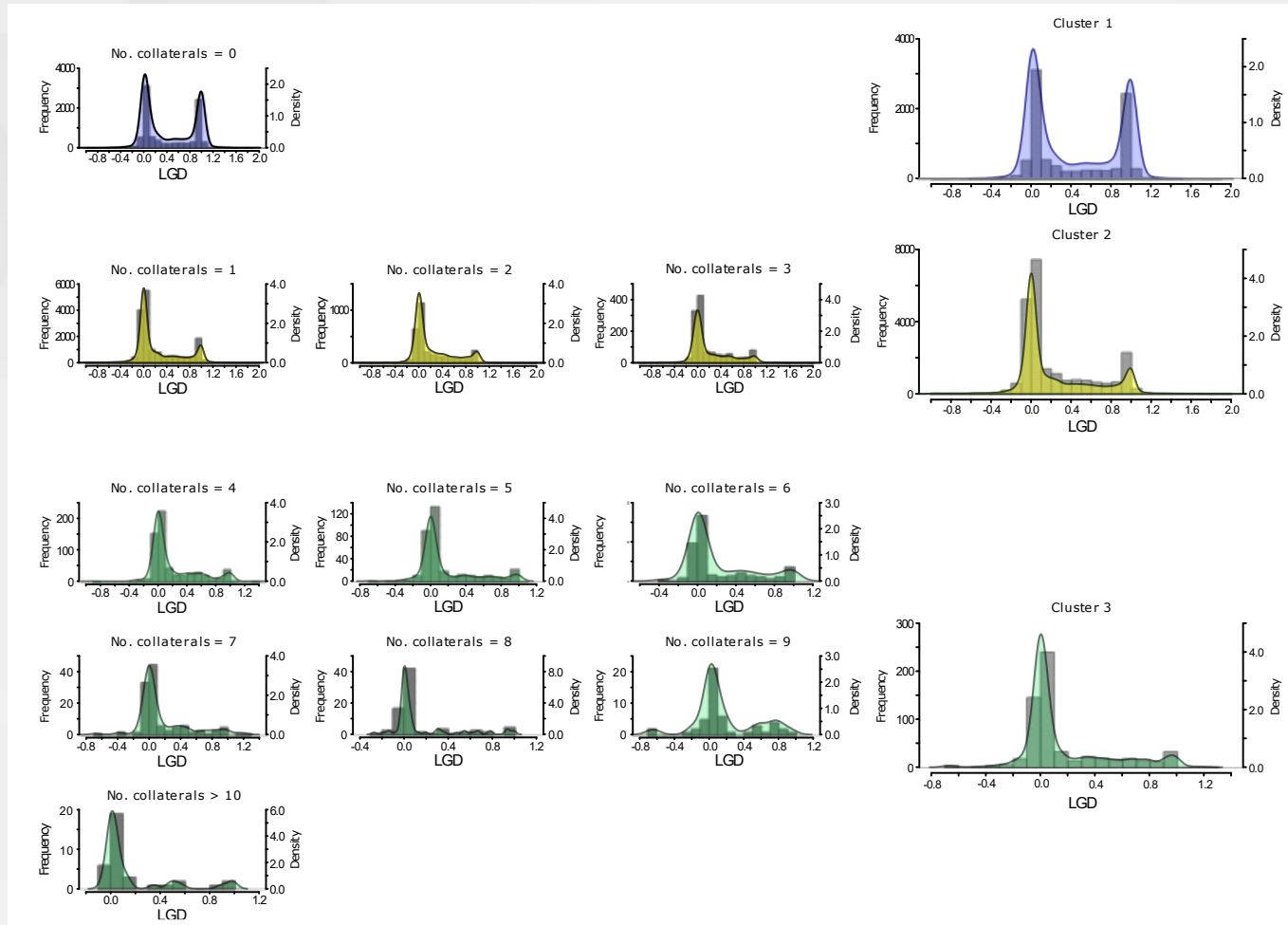


**Table 6** Inclusion of enterprise-specific variables: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1168	0.1182	0.1168	0.1173	0.1172	0.1166	0.1150	0.1181	0.1149	0.1085	0.1153	0.1121	0.1189	0.1163	0.1129	0.1196	0.1142	0.1249	0.1152	0.1168
70/30	0.1140	0.1168	0.1158	0.1157	0.1146	0.1155	0.1148	0.1182	0.1136	0.1062	0.1149	0.1134	0.1158	0.1161	0.1121	0.1177	0.1145	0.1233	0.1144	0.1158
80/20	0.1111	0.1135	0.1143	0.1153	0.1147	0.1156	0.1146	0.1163	0.1160	0.1052	0.1136	0.1173	0.1121	0.1141	0.1136	0.1143	0.1132	0.1204	0.1144	0.1121
90/10	0.1150	0.1089	0.1103	0.1108	0.1124	0.1116	0.1138	0.1121	0.1107	0.1045	0.1131	0.1063	0.1094	0.1127	0.1117	0.1062	0.1085	0.1185	0.1127	0.1107
Mean	0.1142	0.1143	0.1143	0.1148	0.1147	0.1148	0.1145	0.1162	0.1138	<u>0.1061</u>	0.1142	0.1123	0.1140	0.1148	0.1126	0.1144	0.1126	0.1218	0.1142	0.1139
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1279	0.1260	0.1258	0.1268	0.1250	0.1257	0.1242	0.1260	0.1250	0.1238	0.1246	0.1233	0.1270	0.1270	0.1251	0.1277	0.1268	0.1247	0.1254	0.1252
70/30	0.1237	0.1244	0.1242	0.1255	0.1257	0.1252	0.1233	0.1172	0.1204	0.1216	0.1235	0.1210	0.1228	0.1232	0.1257	0.1237	0.1225	0.1227	0.1169	0.1218
80/20	0.1219	0.1217	0.1243	0.1235	0.1246	0.1242	0.1234	0.1173	0.1163	0.1099	0.1236	0.1074	0.1149	0.1184	0.1115	0.1188	0.1090	0.1219	0.1175	0.1182
90/10	0.1192	0.1197	0.1216	0.1191	0.1198	0.1198	0.1177	0.1176	0.1149	0.1088	0.1235	0.1058	0.1062	0.1171	0.1077	0.1179	0.1052	0.1205	0.1148	0.1061
Mean	0.1232	0.1230	0.1240	0.1237	0.1238	0.1237	0.1221	0.1195	0.1192	0.1160	0.1238	<u>0.1144</u>	0.1177	0.1214	0.1175	0.1221	0.1159	0.1225	0.1186	0.1178
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0789	0.0817	0.0835	0.0809	0.0805	0.0805	0.0835	0.0694	0.0664	0.0585	0.0780	0.0589	0.0624	0.0780	0.0592	0.0671	0.0639	0.0681	0.0705	0.0485
70/30	0.0811	0.0815	0.0838	0.0832	0.0791	0.0796	0.0829	0.0679	0.0646	0.0553	0.0764	0.0565	0.0592	0.0778	0.0582	0.0652	0.0649	0.0675	0.0688	0.0472
80/20	0.0769	0.0799	0.0820	0.0769	0.0790	0.0790	0.0818	0.0653	0.0628	0.0532	0.0741	0.0546	0.0575	0.0759	0.0567	0.0646	0.0617	0.0658	0.0663	0.0437
90/10	0.0788	0.0785	0.0812	0.0788	0.0797	0.0791	0.0762	0.0630	0.0621	0.0536	0.0725	0.0525	0.0566	0.0725	0.0563	0.0634	0.0621	0.0661	0.0647	0.0437
Mean	0.0789	0.0804	0.0826	0.0800	0.0796	0.0795	0.0811	0.0664	0.0640	0.0551	0.0753	0.0556	0.0589	0.0760	0.0576	0.0651	0.0631	0.0669	0.0676	<u>0.0458</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit.

# Robustness check: Clustering based on collaterals



# Robustness check: Clustering based on collaterals

**Table 7** Clustering based on loan-specific variable: Out-of-sample estimation accuracies (*MSE*)

Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Cluster 1: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1366	0.1365	0.1366	0.1370	0.1372	0.1374	0.1369	0.1370	0.1357	0.1300	0.1367	0.1348	0.1353	0.1370	0.1353	0.1372	0.1362	0.1448	0.1360	0.1348
70/30	0.1362	0.1357	0.1362	0.1370	0.1369	0.1367	0.1366	0.1366	0.1356	0.1305	0.1357	0.1328	0.1357	0.1368	0.1343	0.1361	0.1359	0.1437	0.1353	0.1345
80/20	0.1355	0.1359	0.1360	0.1363	0.1362	0.1361	0.1360	0.1361	0.1347	0.1262	0.1353	0.1325	0.1340	0.1366	0.1346	0.1364	0.1359	0.1435	0.1344	0.1344
90/10	0.1357	0.1372	0.1361	0.1362	0.1363	0.1363	0.1361	0.1361	0.1354	0.1244	0.1364	0.1313	0.1328	0.1359	0.1331	0.1366	0.1357	0.1438	0.1345	0.1344
Mean	0.1360	0.1363	0.1362	0.1366	0.1367	0.1366	0.1364	0.1364	0.1353	<u>0.1277</u>	0.1360	0.1328	0.1345	0.1366	0.1343	0.1366	0.1359	0.1439	0.1350	0.1345
<b>Cluster 2: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1254	0.1257	0.1258	0.1256	0.1257	0.1256	0.1255	0.1252	0.1235	0.1207	0.1293	0.1194	0.1241	0.1249	0.1216	0.1252	0.1222	0.1253	0.1249	0.1234
70/30	0.1253	0.1254	0.1257	0.1253	0.1253	0.1254	0.1257	0.1247	0.1232	0.1199	0.1302	0.1183	0.1242	0.1248	0.1211	0.1248	0.1225	0.1251	0.1245	0.1229
80/20	0.1252	0.1252	0.1253	0.1253	0.1255	0.1253	0.1249	0.1245	0.1220	0.1193	0.1280	0.1177	0.1239	0.1248	0.1207	0.1247	0.1210	0.1250	0.1239	0.1222
90/10	0.1251	0.1251	0.1254	0.1252	0.1253	0.1253	0.1243	0.1242	0.1213	0.1193	0.1273	0.1161	0.1234	0.1238	0.1196	0.1247	0.1193	0.1248	0.1235	0.1216
Mean	0.1252	0.1253	0.1255	0.1254	0.1254	0.1254	0.1251	0.1246	0.1225	0.1198	0.1287	<u>0.1179</u>	0.1239	0.1246	0.1207	0.1249	0.1213	0.1251	0.1242	0.1225
<b>Cluster 3: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0806	0.0802	0.0824	0.0797	0.0815	0.0815	0.0796	0.0764	0.0758	0.0708	0.0794	0.0735	0.0729	0.0785	0.0754	0.0742	0.0758	0.0754	0.0777	0.0695
70/30	0.0797	0.0790	0.0816	0.0796	0.0797	0.0798	0.0799	0.0757	0.0744	0.0702	0.0777	0.0729	0.0725	0.0787	0.0735	0.0749	0.0760	0.0751	0.0756	0.0674
80/20	0.0778	0.0780	0.0815	0.0802	0.0796	0.0795	0.0779	0.0760	0.0749	0.0695	0.0786	0.0733	0.0723	0.0787	0.0730	0.0738	0.0744	0.0751	0.0759	0.0669
90/10	0.0781	0.0783	0.0816	0.0813	0.0791	0.0800	0.0782	0.0758	0.0748	0.0677	0.0761	0.0726	0.0735	0.0765	0.0727	0.0729	0.0726	0.0756	0.0750	0.0637
Mean	0.0790	0.0788	0.0818	0.0802	0.0800	0.0802	0.0789	0.0760	0.0750	0.0695	0.0780	0.0731	0.0728	0.0781	0.0737	0.0739	0.0747	0.0753	0.0760	<u>0.0669</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit

# Robustness check: Logarithmic transformation C3

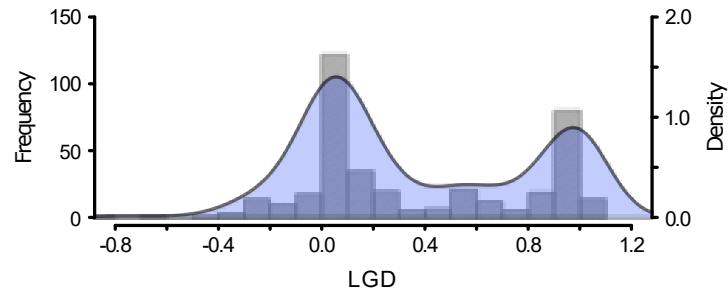
**Table 8** Logarithmic transformation of the positively skewed unimodally distributed LGDs: Out-of-sample estimation accuracies (*MSE*)

Split	<i>OLS</i>	<i>bOLS</i>	<i>LAR</i>	<i>RR</i>	<i>LR</i>	<i>ER</i>	<i>FLR</i>	<i>RT</i>	<i>CIT</i>	<i>RF</i>	<i>ADA</i>	<i>GB</i>	<i>CUB</i>	<i>ANN</i>	<i>SVR</i>	<i>RVR</i>	<i>GAPR</i>	<i>KNN</i>	<i>MARS</i>	<i>FMM</i>
60/40	0.0804	0.0809	0.0848	0.0811	0.0806	0.0817	0.0817	0.0659	0.0653	0.0556	0.0737	0.0618	0.0629	0.0806	0.0581	0.0694	0.0639	0.0761	0.0690	0.0444
70/30	0.0787	0.0776	0.0823	0.0782	0.0780	0.0776	0.0810	0.0652	0.0652	0.0553	0.0742	0.0617	0.0609	0.0768	0.0564	0.0655	0.0628	0.0751	0.0675	0.0434
80/20	0.0775	0.0789	0.0826	0.0779	0.0776	0.0776	0.0813	0.0651	0.0639	0.0549	0.0722	0.0597	0.0627	0.0760	0.0555	0.0655	0.0613	0.0746	0.0675	0.0441
90/10	0.0767	0.0800	0.0824	0.0772	0.0780	0.0775	0.0817	0.0652	0.0622	0.0531	0.0703	0.0618	0.0607	0.0757	0.0542	0.0656	0.0614	0.0698	0.0640	0.0444
Mean	0.0783	0.0793	0.0830	0.0786	0.0786	0.0786	0.0814	0.0653	0.0641	0.0547	0.0726	0.0612	0.0618	0.0773	0.0561	0.0665	0.0623	0.0739	0.0670	<u>0.0441</u>

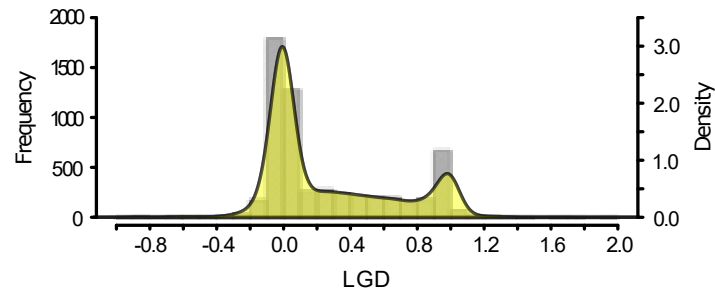
Note. This table reports the *MSE* of the out-of-sample estimation for the considered LGD methods. High values for *MSE* imply a bad fit

# Robustness check: Non-European credit portfolio

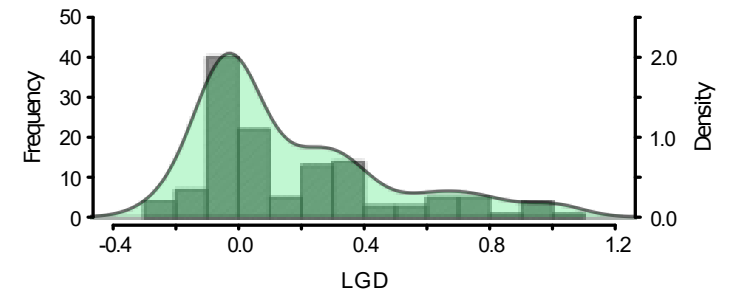
Latin America



North America



Oceania



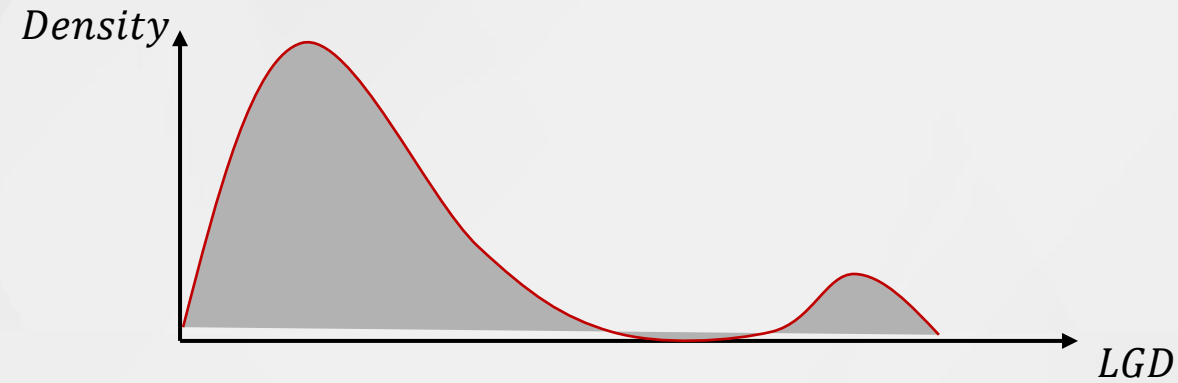
# Robustness check: Non-European credit portfolio

**Table 9** Non-European credit portfolios: Out-of-sample estimation accuracies (*MSE*)

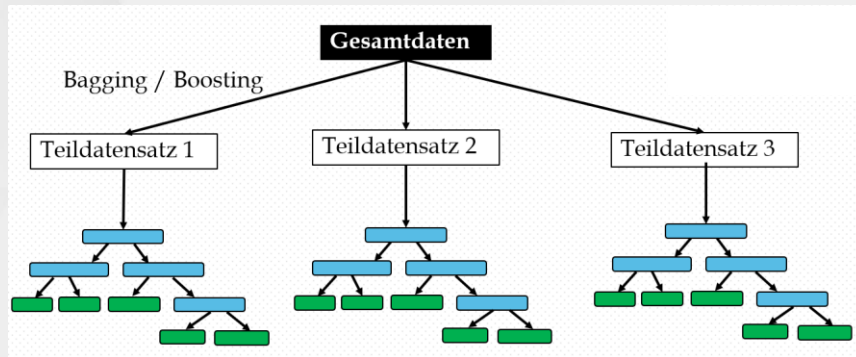
Split	OLS	bOLS	LAR	RR	LR	ER	FLR	RT	CIT	RF	ADA	GB	CUB	ANN	SVR	RVR	GAPR	KNN	MARS	FMM
<b>Latin American credit portfolio: (nearly) symmetric bimodal LGD distribution</b>																				
60/40	0.1403	0.1465	0.1391	0.1390	0.1476	0.1478	0.1383	0.1458	0.1414	0.1256	0.1429	0.1491	0.1472	0.1448	0.1429	0.1421	0.1258	0.1444	0.1461	0.1338
70/30	0.1388	0.1467	0.1388	0.1371	0.1468	0.1458	0.1367	0.1421	0.1395	0.1110	0.1395	0.1126	0.1112	0.1434	0.1415	0.1384	0.1035	0.1433	0.1459	0.1298
80/20	0.1344	0.1449	0.1366	0.1374	0.1444	0.1450	0.1361	0.1440	0.1374	0.1072	0.1375	0.1196	0.1183	0.1409	0.1413	0.1368	0.1331	0.1443	0.1449	0.1267
90/10	0.1353	0.1438	0.1378	0.1303	0.1434	0.1443	0.1351	0.1420	0.1260	0.1081	0.1378	0.1123	0.1106	0.1398	0.1362	0.1263	0.1082	0.1373	0.1406	0.1098
Mean	0.1372	0.1455	0.1381	0.1359	0.1455	0.1457	0.1366	0.1435	0.1361	<u>0.1130</u>	0.1394	0.1234	0.1218	0.1422	0.1405	0.1359	0.1176	0.1423	0.1444	0.1250
<b>North American credit portfolio: asymmetric (positively skewed) bimodal LGD distribution</b>																				
60/40	0.1363	0.1367	0.1363	0.1363	0.1402	0.1381	0.1334	0.1391	0.1390	0.1265	0.1438	0.1272	0.1334	0.1381	0.1346	0.1384	0.1313	0.1387	0.1371	0.1342
70/30	0.1361	0.1362	0.1361	0.1361	0.1395	0.1377	0.1331	0.1387	0.1358	0.1251	0.1429	0.1249	0.1287	0.1378	0.1339	0.1379	0.1286	0.1382	0.1305	0.1337
80/20	0.1344	0.1336	0.1345	0.1344	0.1388	0.1371	0.1313	0.1351	0.1362	0.1228	0.1411	0.1179	0.1241	0.1348	0.1334	0.1359	0.1264	0.1380	0.1284	0.1268
90/10	0.1289	0.1284	0.1289	0.1289	0.1348	0.1365	0.1258	0.1332	0.1290	0.1198	0.1393	0.1185	0.1262	0.1351	0.1289	0.1351	0.1253	0.1342	0.1250	0.1248
Mean	0.1339	0.1337	0.1340	0.1339	0.1383	0.1373	0.1309	0.1365	0.1350	0.1236	0.1418	<u>0.1221</u>	0.1281	0.1364	0.1327	0.1368	0.1279	0.1373	0.1303	0.1299
<b>Oceanian credit portfolio: (positively skewed) unimodal LGD distribution</b>																				
60/40	0.0992	0.1019	0.1244	0.0994	0.0943	0.0927	0.1010	0.0951	0.1074	0.0764	0.0974	0.0893	0.0958	0.1062	0.0936	0.1094	0.0847	0.1263	0.1080	0.0651
70/30	0.1065	0.0951	0.1236	0.0951	0.0922	0.0933	0.0971	0.0918	0.1065	0.0801	0.0943	0.0888	0.0922	0.1058	0.0923	0.1077	0.0828	0.1254	0.1083	0.0560
80/20	0.1187	0.1007	0.1216	0.0977	0.0916	0.0922	0.1040	0.0919	0.1075	0.0748	0.0941	0.0885	0.0937	0.1054	0.0914	0.1016	0.0808	0.1262	0.1075	0.0514
90/10	0.1117	0.0964	0.1182	0.0959	0.0908	0.0914	0.0988	0.0904	0.1071	0.0764	0.0972	0.0882	0.0922	0.1014	0.0915	0.1023	0.0727	0.1241	0.1030	0.0410
Mean	0.1090	0.0985	0.1220	0.0970	0.0922	0.0924	0.1002	0.0923	0.1071	0.0769	0.0957	0.0887	0.0935	0.1047	0.0922	0.1052	0.0803	0.1255	0.1067	<u>0.0534</u>

Note. This table reports the *MSE* of the out-of-sample estimation for the Latin American, North American, and Oceanian credit portfolios. High values for *MSE* imply a bad fit

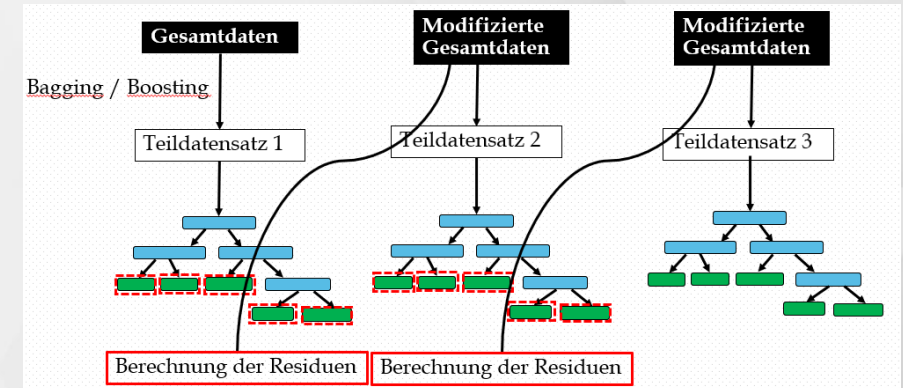
# Random Forest vs. Boosting



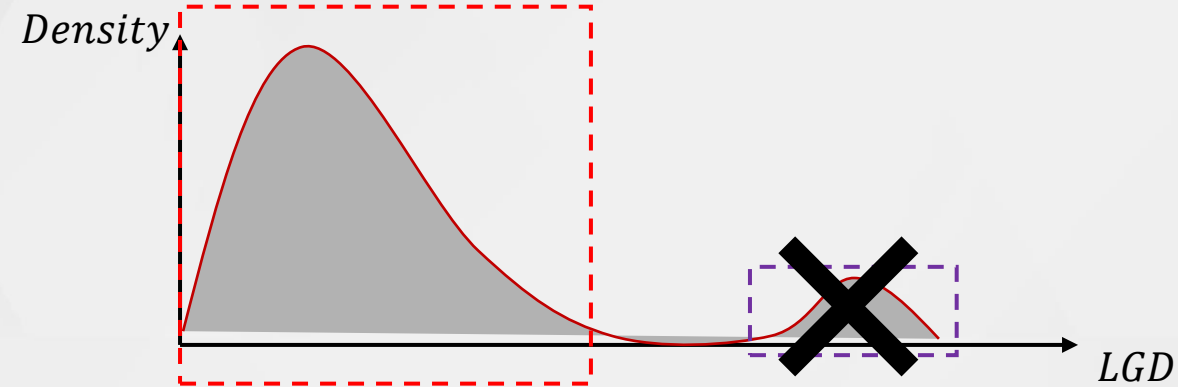
Random Forest



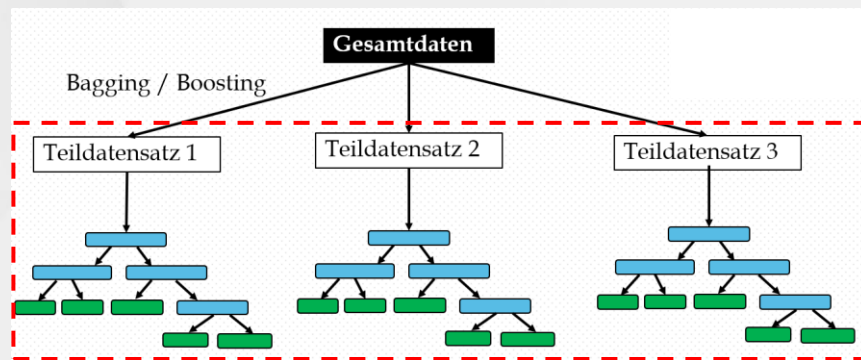
Gradient Boosting



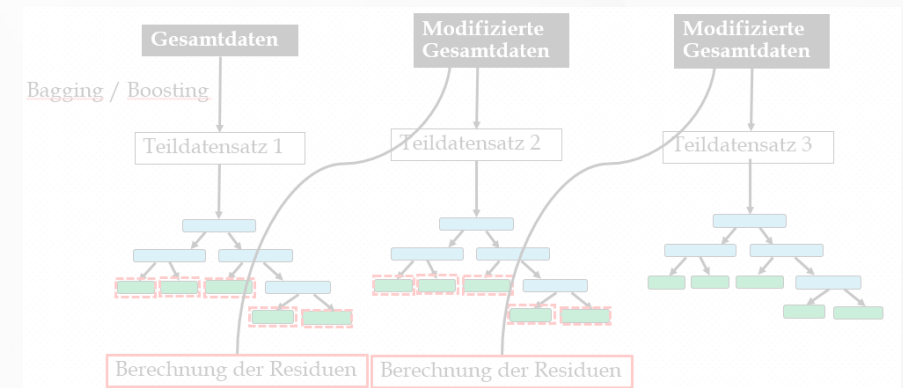
# Random Forest vs. Boosting



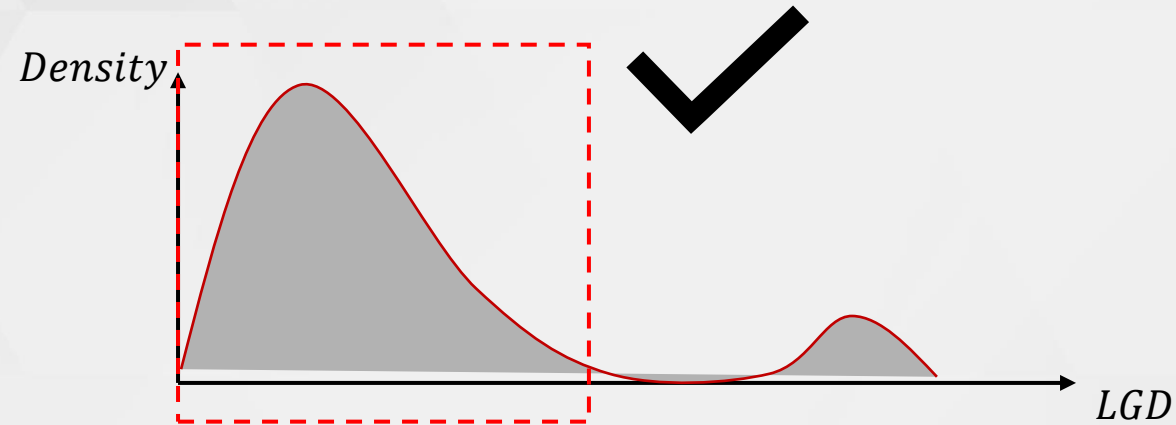
Random Forest



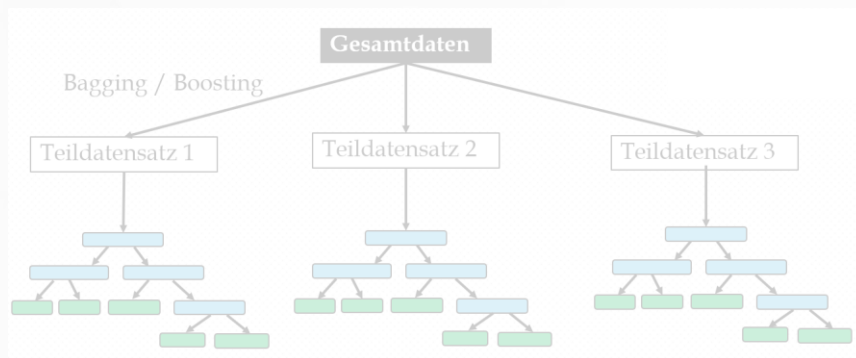
Gradient Boosting



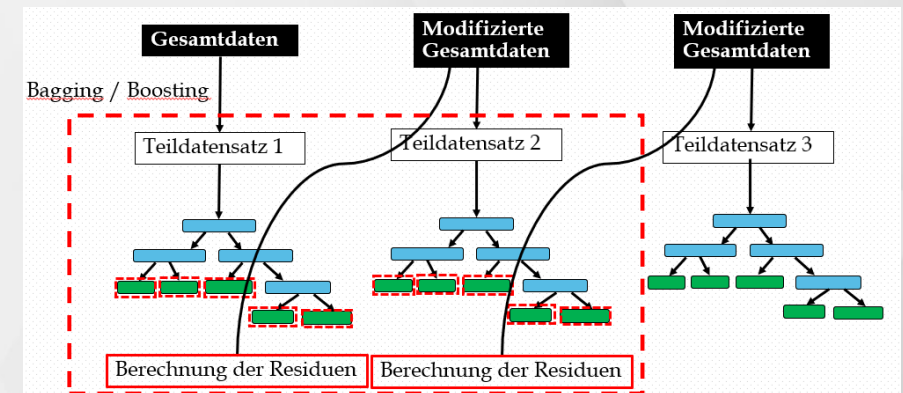
# Random Forest vs. Boosting



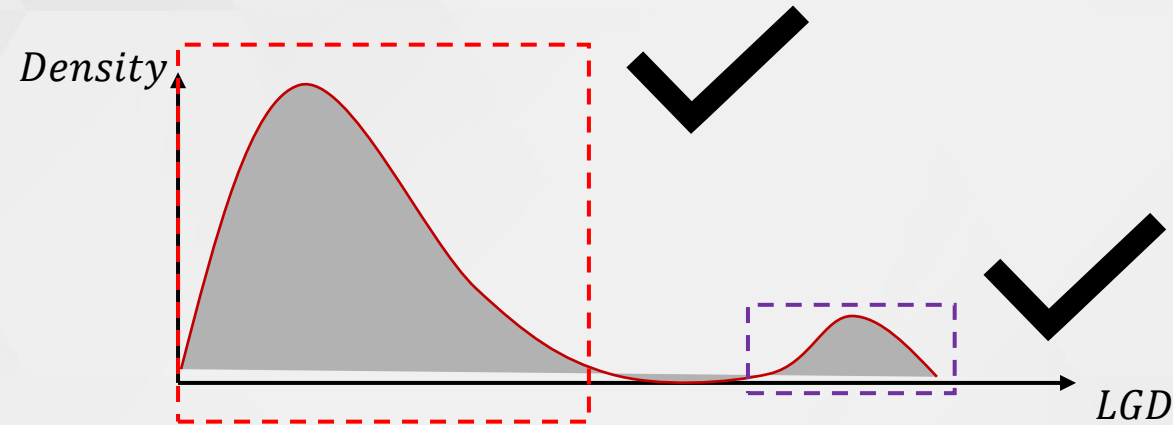
## Random Forest



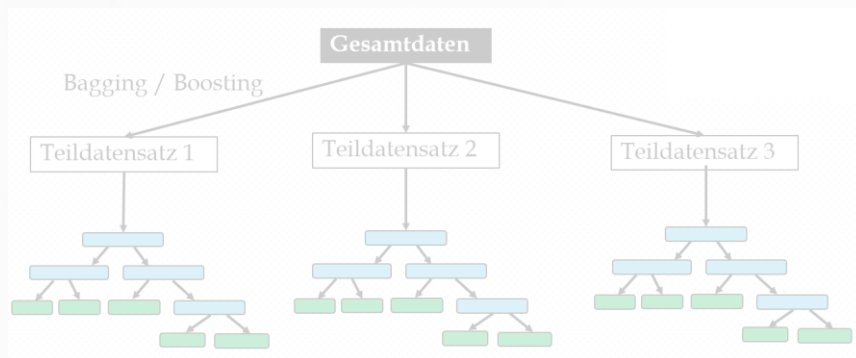
## Gradient Boosting



# Random Forest vs. Boosting



Random Forest



Gradient Boosting

