

A zero-adjusted gamma model for estimating loss given default on residential mortgage loans

Edward Tong, Christophe Mues, Lyn Thomas
e.tong@soton.ac.uk

Credit Scoring and Credit Control XII conference
Edinburgh, August 24-26 2011

Loss given default (LGD)

- Basel II - requirement for Internal Ratings Based (IRB) Advanced approach for calculating minimum capital requirements
- LGD defined as the proportion of the loan lost in the event of default

$$\text{Gross LGD} = \frac{\text{Total Loss}}{\text{Exposure At Default}}$$

- This study: LGD for residential mortgage loans

Some mortgage LGD models

- LGD is derived using a combination of probability of repossession and haircut models (Lucas, 2006; Leow & Mues 2011)
- Haircut model with quantile regression (Somers & Whitaker, 2007)
- Linear regression used to directly model LGD with high loan to value (Qi & Yang, 2009)

Research objectives

- To directly model the loss amount to derive an estimate of LGD
- Explore potential non-linearity between predictor variables and loss amount response variable
- Compare performance of loss amount model with a well known approach used in industry

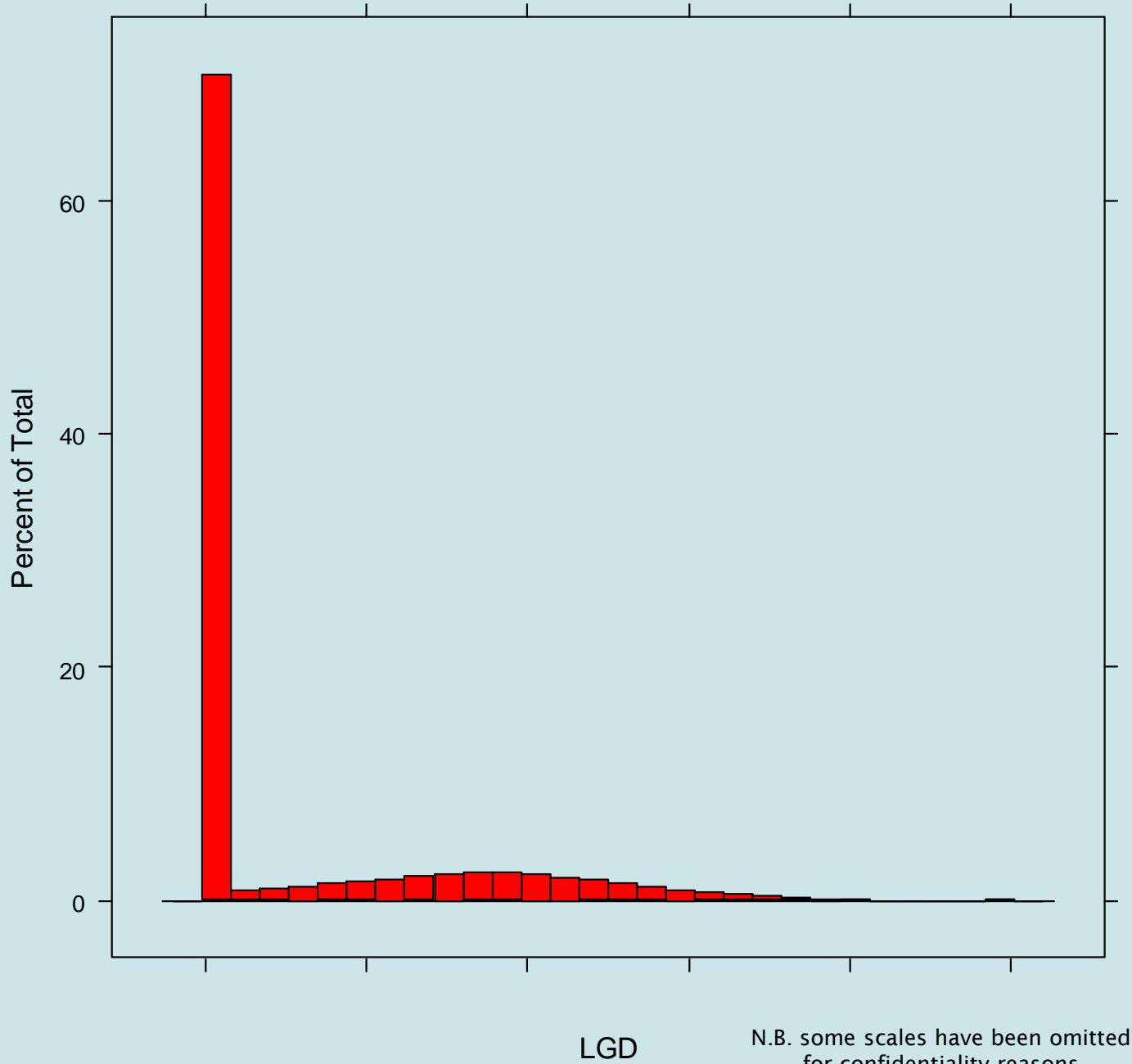
Data

- UK bank
- Residential mortgage portfolio
- 13 years of data (1988 to 2000)
- All observations are defaulted mortgages
- >113,000 defaults in total sample

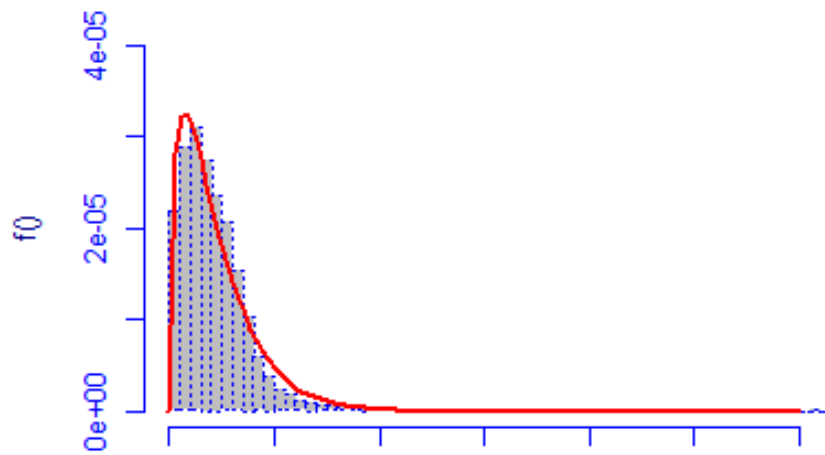
Data (cont'd)

- Response variable – loss amount
- 21 application and behavioural variables including - loan balance at default, indexed valuation of property at default, time on books, loan-to-value (LTV), debt-to-value (DTV), previous default indicator, loan term, geographical region, HPI growth rate at default quarter

Distribution of LGD

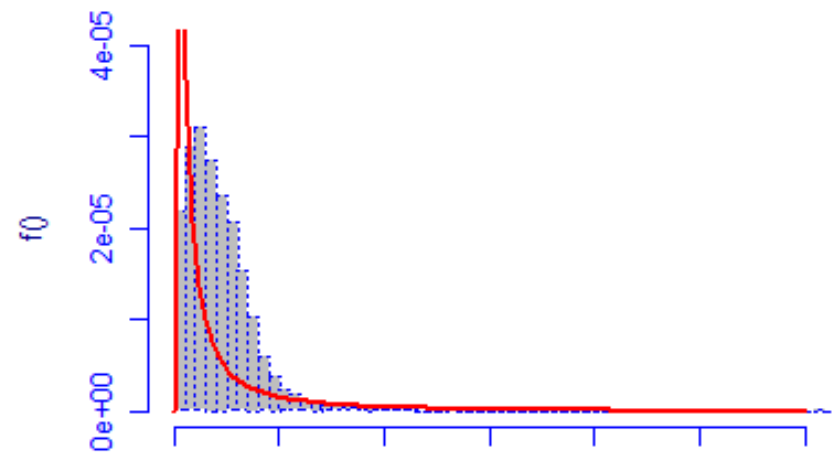


Fitted gamma distribution (training set)



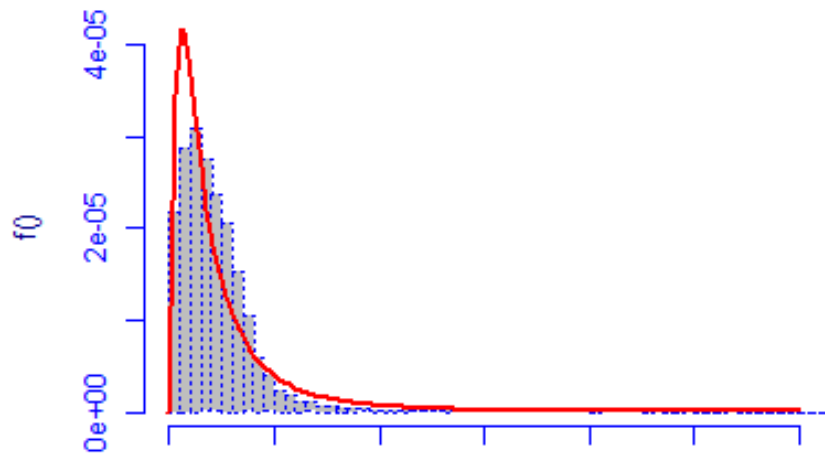
Non-zero loss amount (GBP)

Fitted inverse Gaussian distribution (training set)



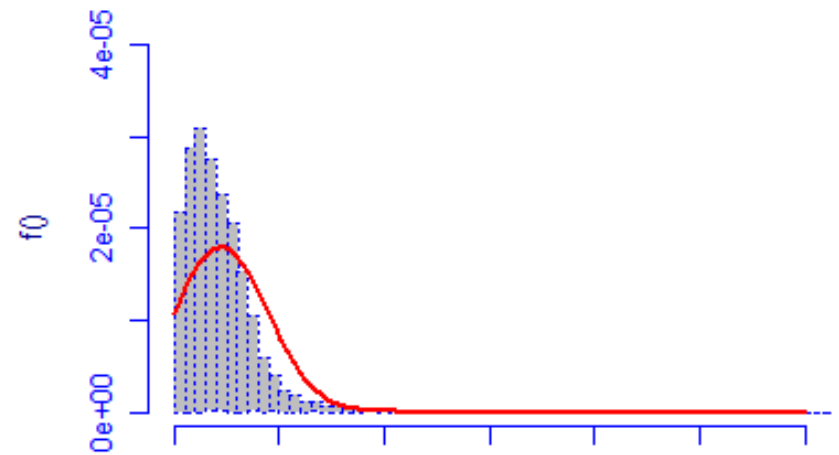
Non-zero loss amount (GBP)

Fitted log normal distribution (training set)



Non-zero loss amount (GBP)

Fitted normal distribution (training set)



Non-zero loss amount (GBP)

Zero-adjusted gamma distribution

The probability function of the ZAGA is defined by

$$f_Y(y | \mu, \sigma, \pi) = \begin{cases} \pi & \text{if } y = 0 \\ (1 - \pi) \left[\frac{1}{(\sigma^2 \mu)^{1/\sigma^2}} \frac{y^{\frac{1}{\sigma^2} - 1} e^{-y/(\sigma^2 \mu)}}{\Gamma(1/\sigma^2)} \right] & \text{if } y > 0 \end{cases}$$

for $0 \leq y < \infty$, $0 < \pi < 1$, $\mu > 0$, $\sigma > 0$

where μ denotes mean, σ scale,

π probability of zero loss

$$E(Y) = (1 - \pi) \mu \text{ and } Var(Y) = (1 - \pi) \mu^2 (\pi + \sigma^2)$$

Generalized Additive Model for Location, Scale & Shape

- GAMLSS (Rigby & Stasinopoulos, 2005, 2007) implemented in `gamlss` package in R
- General framework for fitting regression type models
- Response variable $y \sim D(y \mid \mu, \sigma, \nu, \tau)$ where $D()$ can be any distribution (over 50 different types including highly skew and kurtotic continuous and discrete distributions)

ZAGA model setup

$$\log(\mu) = \eta_1 = X_1\beta_1 + \sum_{j=1}^{J_1} h_{j1}(x_{j1})$$

$$\log(\sigma) = \eta_2 = X_2\beta_2 + \sum_{j=1}^{J_2} h_{j2}(x_{j2})$$

$$\text{logit}(\pi) = \eta_3 = X_3\beta_3 + \sum_{j=1}^{J_3} h_{j3}(x_{j3})$$

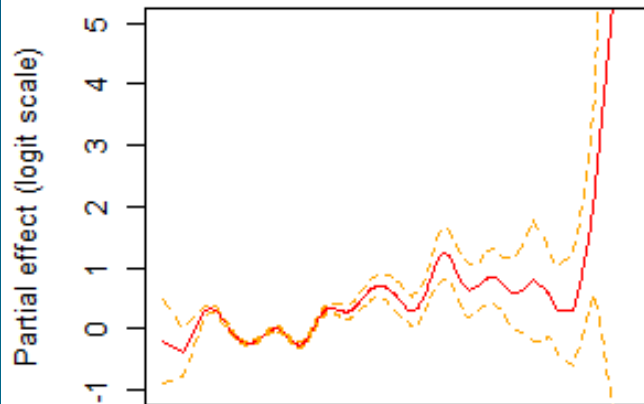
where $X_k\beta_k$ denote parametric linear terms,

$h_{jk}(x_{jk})$ denote additive smoothers

ZAGA model development

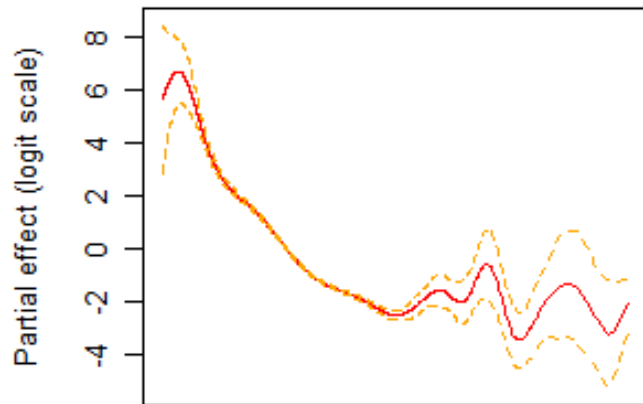
- Separate model components estimated for μ , σ and π components
- Developed with stepwise selection based on Akaike Information Criteria (AIC)
- Continuous variables fitted with smoothers based on penalized *B*-splines (Eilers & Marx, 1996)

Occurrence of zero loss



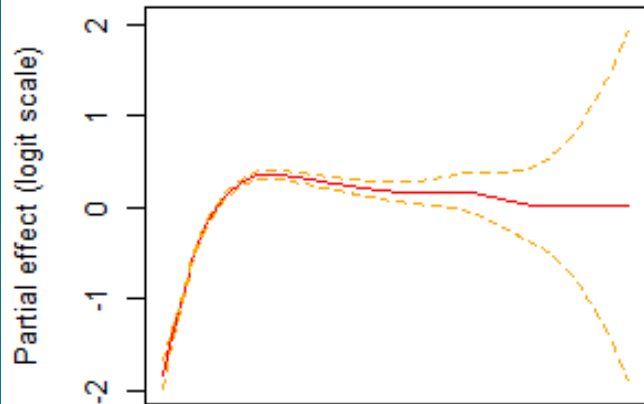
HPI growth rate at default (%)

Occurrence of zero loss



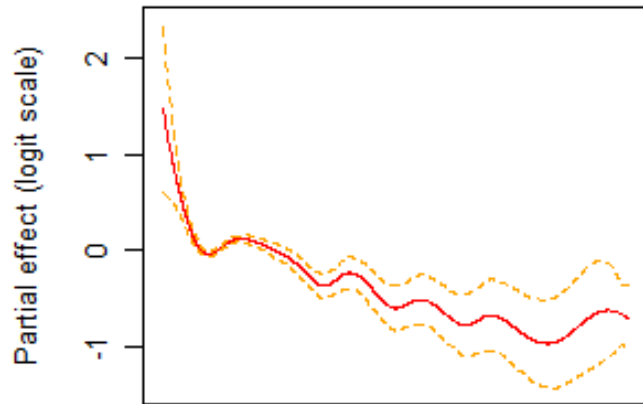
Debt-to-value ratio (DTV)

Occurrence of zero loss



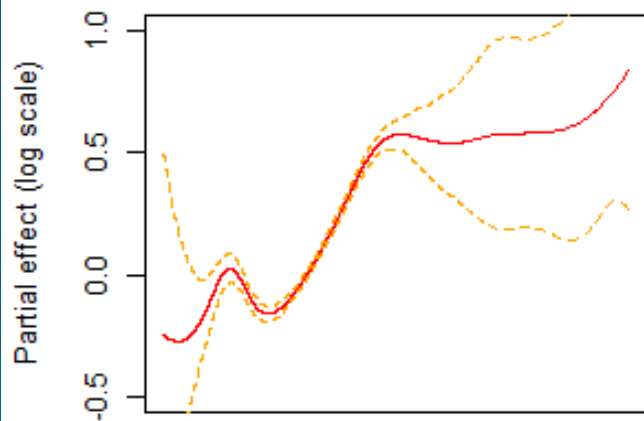
Time on books (years)

Occurrence of zero loss



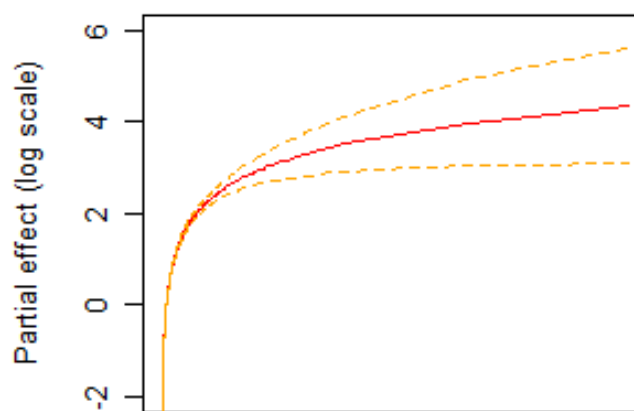
Valuation at default to average valuation for region

Mean of non-zero loss amount



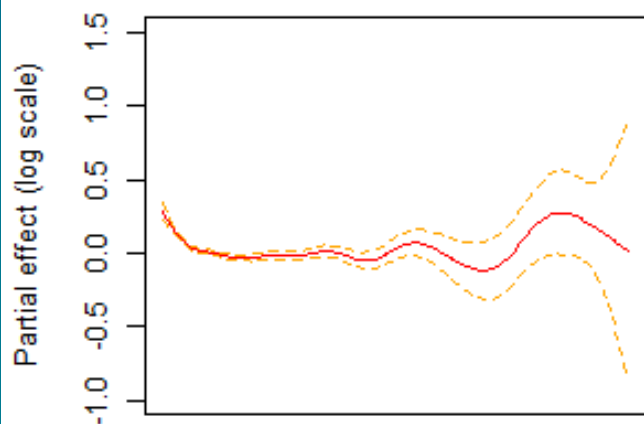
Debt-to-value ratio (DTV)

Mean of non-zero loss amount



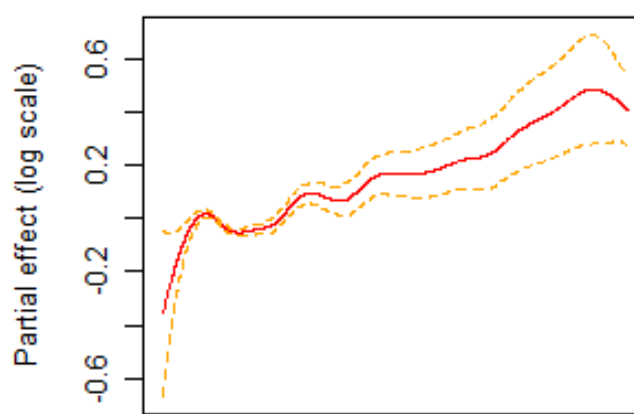
Final loan balance (GBP)

Mean of non-zero loss amount



Time on books (years)

Mean of non-zero loss amount



Valuation at default to average valuation for region

Comparison to linear regression (OLS) with beta transformation

- Based on LossCalc (Gupton & Stein, 2005)

$$Z = \Phi^{-1} \left[\text{Beta}(LGD, \alpha, \beta, \varepsilon) \right]$$

- Assume LGD is beta distributed, estimate α and β parameters from LGD, after adding ε
- Cumulative probabilities are computed using α and β
- Use inverse standard normal to transform from $(0, 1)$ to $(-\infty, \infty)$ and run OLS on Z

OLS-beta (cont'd)

- OLS fitted using polynomial regression
- Variables selected through stepwise regression with backward elimination based on Akaike Information Criteria (AIC)
- Sensitivity to ε ; an adjustment amount for zero losses is necessary because the inverse normal and beta transform is undefined at zero (Qi & Zhao, 2011)

OLS-beta: Sensitivity to ε

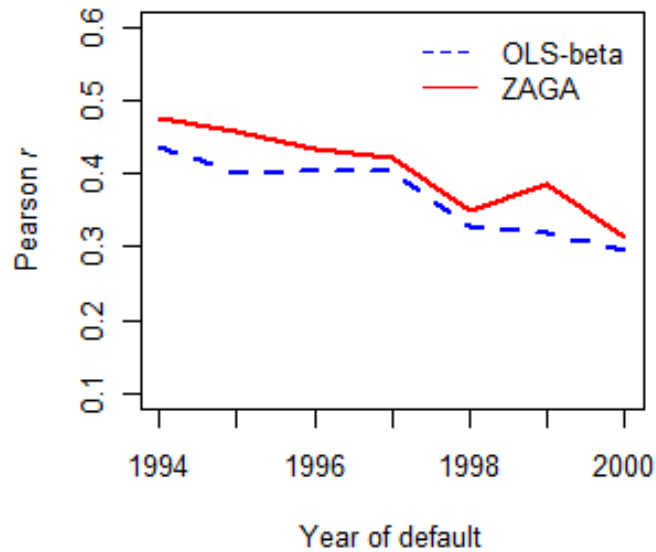
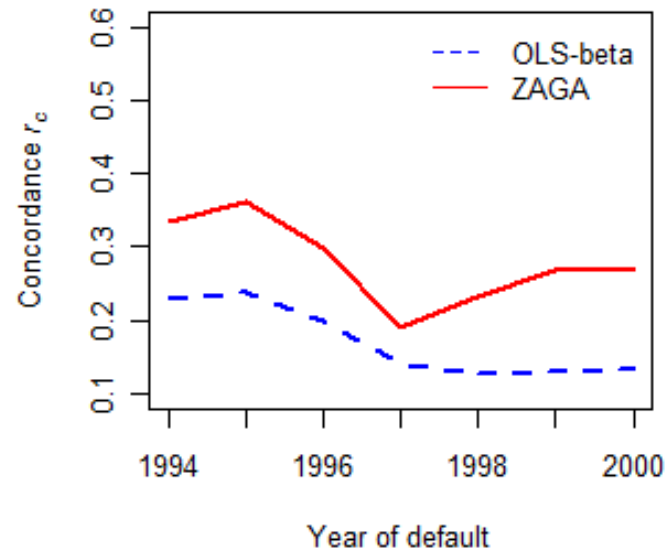
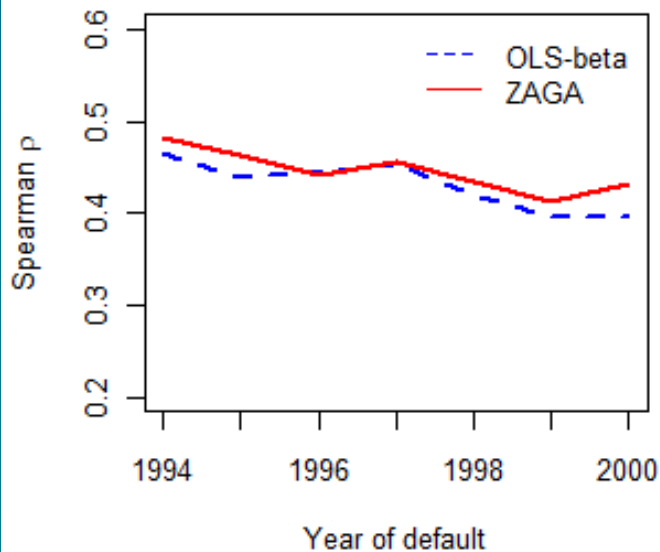
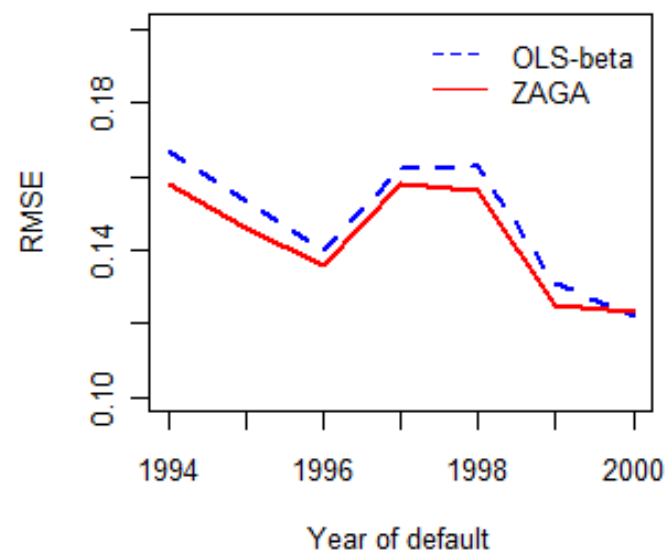
Training years	ε	R^2	Bootstrap SE	RMSE	Bootstrap SE
1988-1994	1.00E-11	0.323	0.003	1.441	0.003
	0.0001	0.333	0.003	0.843	0.002
	0.0005	0.334	0.003	0.756	0.002
	0.001	0.335	0.003	0.717	0.002
	0.005	0.336	0.003	0.629	0.002
	0.01	0.336	0.003	0.599	0.002
	0.05	0.328	0.003	0.599	0.002
	0.06	0.324	0.003	0.610	0.003
1988-1995	1.00E-11	0.312	0.002	1.415	0.003
	0.0001	0.321	0.002	0.829	0.002
	0.0005	0.322	0.002	0.743	0.002
	0.001	0.323	0.002	0.705	0.002
	0.005	0.324	0.002	0.619	0.002
	0.01	0.324	0.003	0.591	0.002
	0.05	0.315	0.003	0.596	0.002
	0.06	0.312	0.003	0.608	0.002

Validation: Discrimination & Calibration

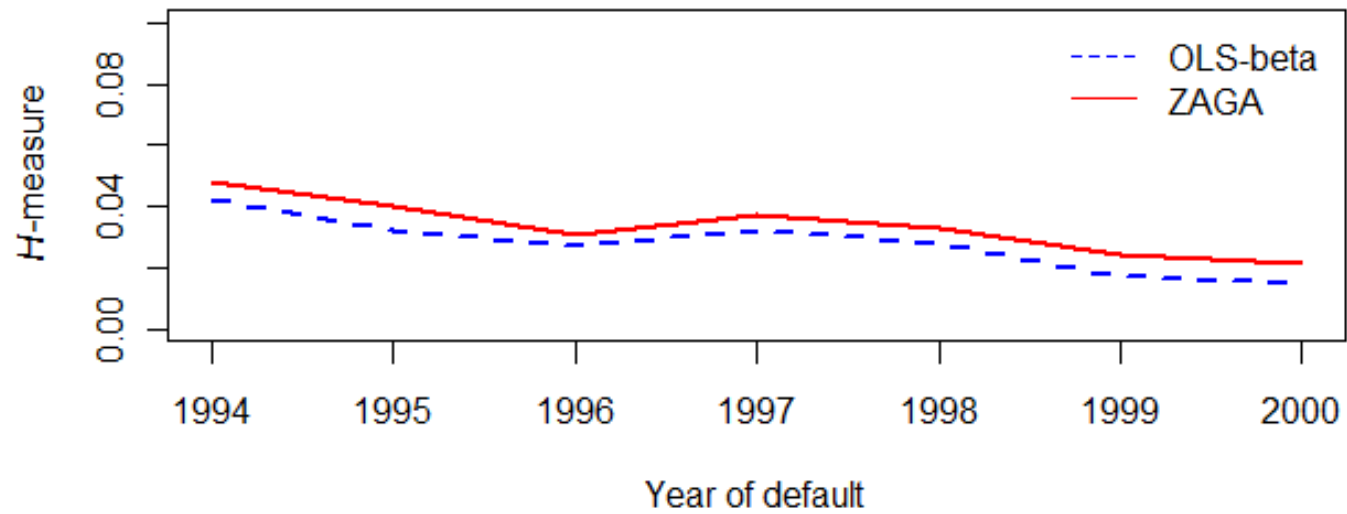
- Pearson r
- Concordance r_c (Lin, 2000)
- Spearman ρ
- RMSE
- AUC (higher vs. lower than average LGD)
- H -measure (Hand, 2009)

Validation: Walk-forward testing

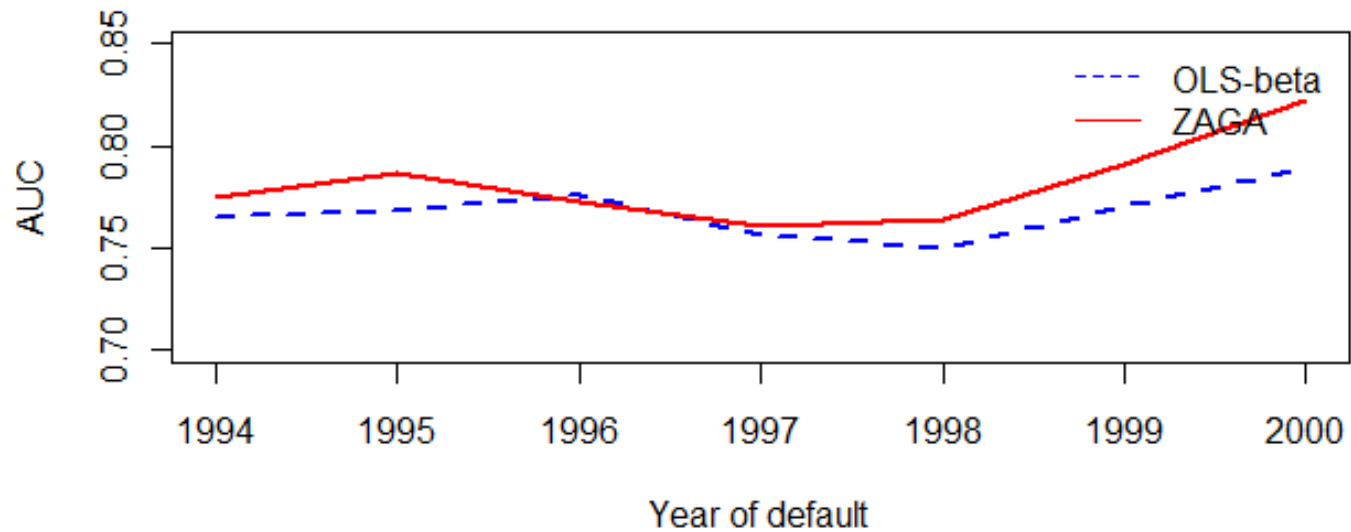
- Special case of cross validation
- Out-of-sample and out-of-time testing
- Train model with first 6 years of data, validate on 7th year
- Next, train model with first 7 years of data, validate on 8th year
- Process is repeated until 7 years of validation folds are obtained

Pearson r **Concordance r_c** **Spearman ρ** **RMSE**

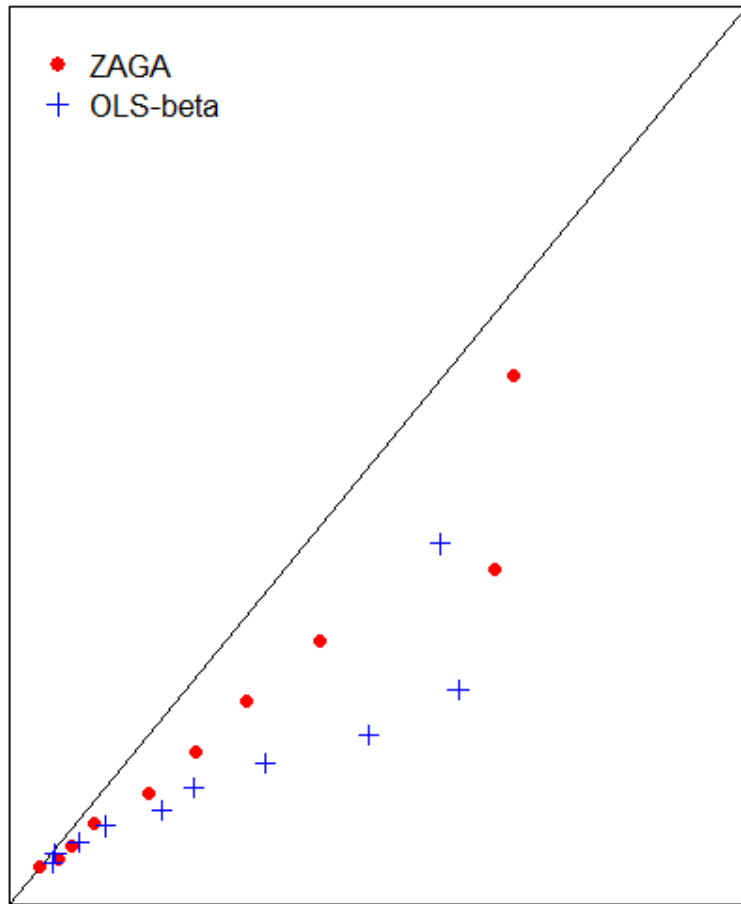
H-measure



AUC

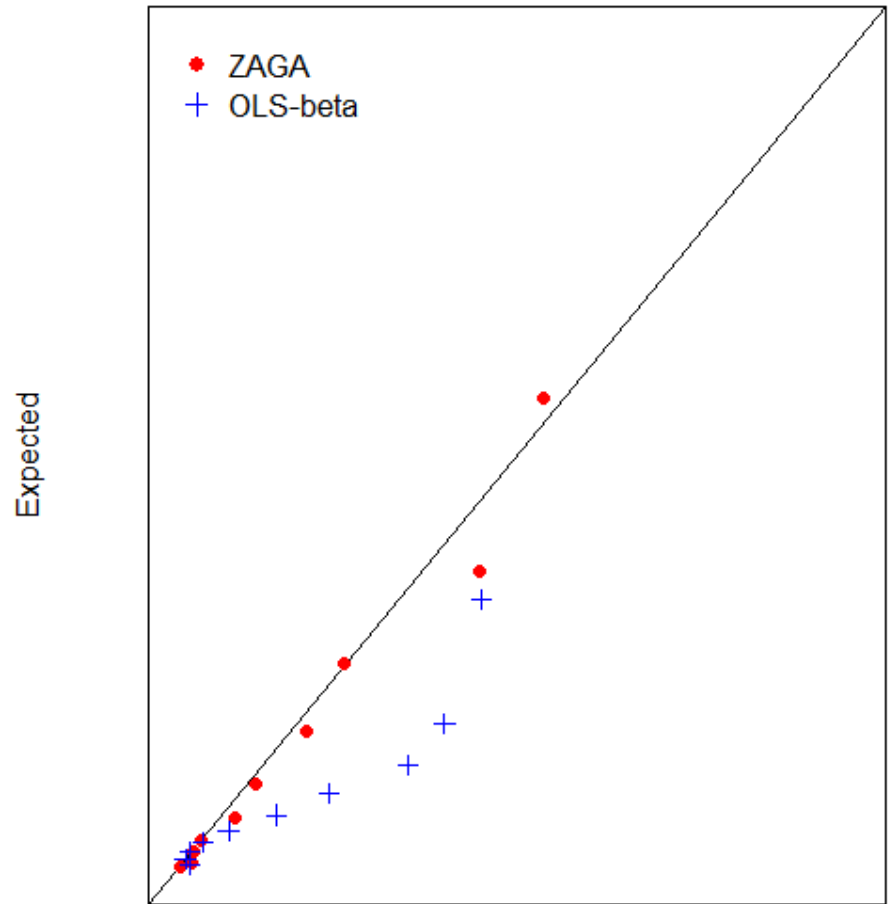


Average LGD by decile bands: 1998



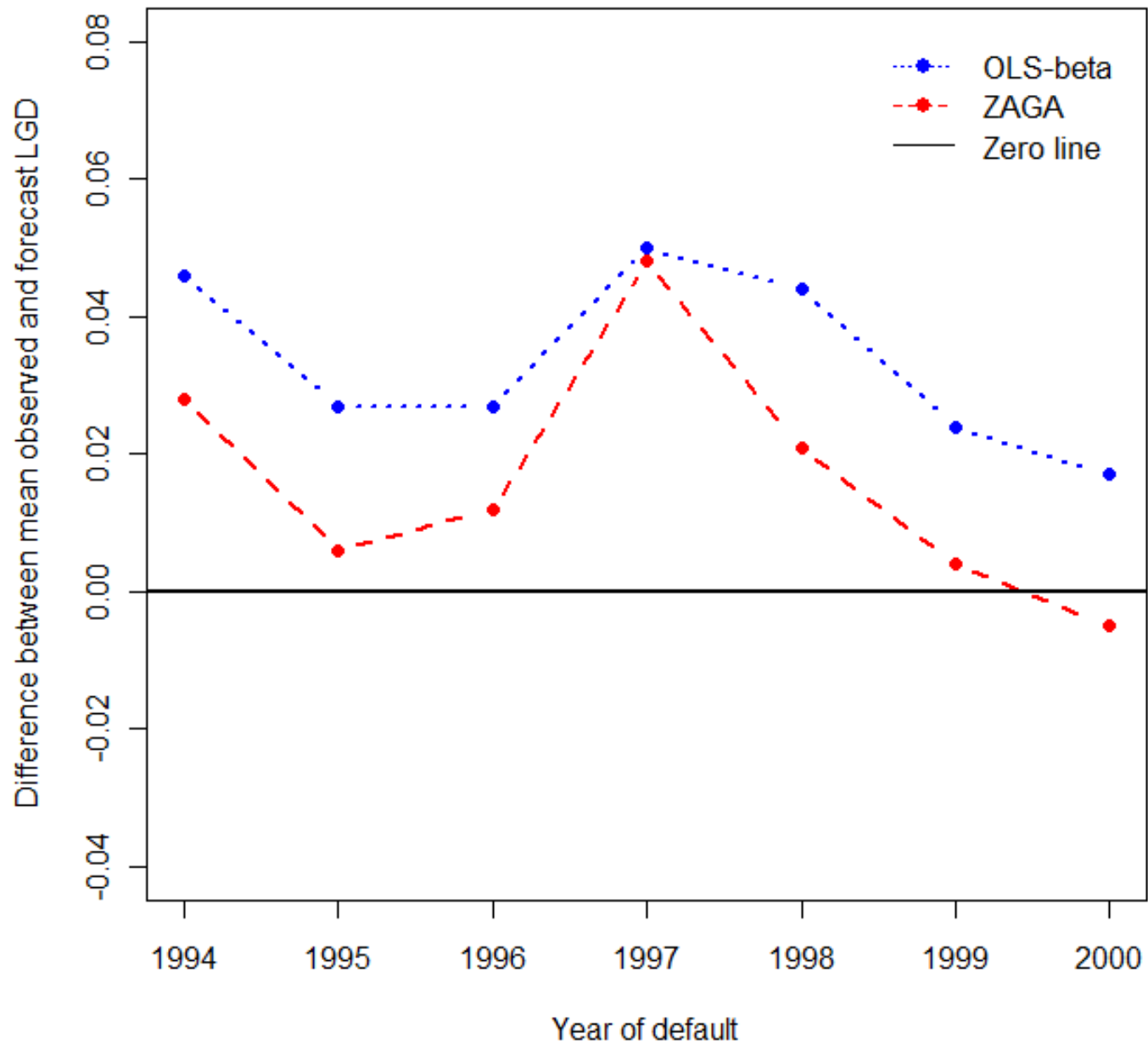
Observed

Average LGD by decile bands: 1999



Observed

Difference between mean observed and forecast LGD



Conclusions

- Modelling the loss amount directly can produce competitively predictive LGD models
- ZAGA model accommodates non-linearity between loss amount and predictors without a black-box approach
- ZAGA mixture approach estimates factors that predict probability of loss and factors that influence the loss amount

Future research

- Potential to improve predictive performance by inclusion of further macroeconomic variables
- Downturn LGD – varying HPI and GDP growth for downturn estimates
- Estimating total losses at a portfolio-level with ZAGA similar to what has been done in insurance policy claims (Heller et al., 2007)

References

- Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2), 89-102.
- Gupton, G. M. & Stein, R. M. (2005). LossCalc v2: Dynamic prediction of LGD. *Moody's KMV*.
- Hand, D. (2009). Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Machine Learning*, 77(1), 103-123.
- Heller G. Z., Stasinopoulos M.D., Rigby R. A. and de Jong P. (2007) Mean and dispersion modeling for policy claims costs. *Scandinavian Actuarial Journal*, 4, 281-292.
- Leow, M. & Mues, C. (2011). Predicting loss given default (LGD) for residential mortgage loans: A two-stage model and empirical evidence for UK bank data. *International Journal of Forecasting*, In Press, Corrected Proof.
- Lin, L. I. K. (2000). A note on the concordance correlation coefficient. *Biometrics*, 56(1), 324-325.
- Lucas, A. (2006). Basel II problem solving. *Conference on Basel II and credit risk modelling in consumer lending*, Southampton, UK.
- Qi, M., & Yang, X. (2009). Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Finance*, 33(5), 788-799.
- Qi, M. & Zhao, X. (2011). Comparison of modeling methods for Loss Given Default. *Journal of Banking & Finance*, In Press, Corrected Proof.
- Rigby, R. A. and Stasinopoulos D. M. (2005). Generalized Additive Models for Location, Scale and Shape. *Applied Statistics*, 54, 507-554.
- Rigby, R. A. & Stasinopoulos, D. M. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, 23.
- Somers, M., & Whittaker, J. (2007). Quantile regression for modelling distributions of profit and loss. *European Journal of Operational Research*, 183, 1477-1487.