

2006



© Institute of Actuaries of Australia

Financial Services Forum

Expanding Our Horizons

Mixture Models for Understanding Dependencies of Insurance and Credit Risks

Emiliano A. Valdez, PhD, FSA, FIAA
University of New South Wales

Brighton-Le-Sands, Sydney
11-12 May 2006



Outline of talk

- Bernoulli mixture models for:
 - credit risk dependencies
 - claim dependencies in an insurance portfolio
- Some early models of dependence
- What are mixture models?
 - pros and cons
- Covariate information
- Data calibration



Some notation

- Consider a portfolio of n insurance policies during some well-defined fixed reference period.
- Random vector of claim indicators: $\mathbf{I} = (I_1, I_2, \dots, I_n)'$
- Each policy k , $k = 1, \dots, n$, comes with a random variable giving an indication for claims:

$$I_k = \begin{cases} 0, & \text{if no claim occurs} \\ 1, & \text{if a claim occurs} \end{cases} .$$

- Joint probability function:
 $p(\mathbf{I}) = P(I_1 = i_1, \dots, I_n = i_n)$ for $i_k \in \{0, 1\}$, $k = 1, \dots, n$.
- Mean vector: $\mathbf{q} = (q_1, q_2, \dots, q_n)'$ where each $q_k = P(I_k = 1)$ is the probability of a claim.



Similar set-up for credit risk portfolios

- One can also think of this portfolio of policies as:
 - a portfolio of credit risks
 - with each default indicator

$$I_k = \begin{cases} 0, & \text{if no default occurs} \\ 1, & \text{if a default occurs} \end{cases} .$$

- The primary interest is to model the joint probability function:

$$p(\mathbf{I}) = P(I_1 = i_1, \dots, I_n = i_n)$$

where in the most general sense the Bernoulli random variables I_k 's are dependent random variables.



Multivariate bernoulli

- Random vector \mathbf{I} is the most general possible case of a multivariate Bernoulli.
- I_k has a Bernoulli(q_k) where q_k is computed

$$q_k = \sum_{i_1=0}^1 \cdots \sum_{i_{k-1}=0}^1 \sum_{i_{k+1}=0}^1 \cdots \sum_{i_n=0}^1 p(\mathbf{I}).$$

- A special case of a multinomial distribution with a total of $2^n - 1$ parameters.
- Sum is no longer a Binomial but is rather a correlated binomial random variable.
- Not often useful for many practical applications because of the large numbers of parameters to estimate.



Some models of dependence

- The Frechet classes - DHAENE AND GOOVAERTS (1997), MULLER (1997)
- Common shock models - MARCEAU, ET AL. (1999)
 - global shock
 - local shock
- Archimedean copula models - GENEST ET AL. (2000), FREY AND MCNEIL (2003)
- Latent variable models and copulas
- Time-until-default (or time-until-claim) and common shock models - LINDSKOG AND MCNEIL (2003)



Dependence through mixing

- See MCNEIL, FREY AND EMBRECHTS (2005) in the context of credit risks.
- Conditional on unobservable $\mathbf{Z} = (Z_1, \dots, Z_p)'$,

$$I_{k|\mathbf{Z}} = I_k | \mathbf{Z} \text{ for all } k \in \{1, 2, \dots, n\}$$

are independently, but not necessarily identically distributed.

- Exchangeable in case identical.
- Generally the dimension p of \mathbf{Z} is smaller than n .
- Assume there are $Q_k : \mathbf{R}^p \rightarrow [0, 1]$ for $k \in \{1, 2, \dots, n\}$ such that

$$P(I_k = 1 | \mathbf{Z}) = p_{k|\mathbf{Z}}(1 | \mathbf{z}) = Q_k(\mathbf{Z}).$$



Marginal and joint probabilities

- Unconditional marginal probability:

$$q_k = P(I_k = 1) = \int Q_k(\mathbf{Z}) dF_{\mathbf{Z}}(\mathbf{z}) = E_{\mathbf{Z}}[Q_k(\mathbf{Z})]$$

- Joint (conditional) probability:

$$P(I_1 = i_1, \dots, I_n = i_n | \mathbf{Z}) = \prod_{k=1}^n p_{k|\mathbf{Z}}(i_k | \mathbf{z}) = \prod_{k=1}^n Q_k(\mathbf{Z})^{i_k} (1 - Q_k(\mathbf{Z}))^{1-i_k}$$

- Joint (unconditional) probability:

$$\begin{aligned} p_{\mathbf{I}}(i_1, \dots, i_n) &= P(I_1 = i_1, \dots, I_n = i_n) \\ &= \int \left[\prod_{k=1}^n Q_k(\mathbf{Z})^{i_k} (1 - Q_k(\mathbf{Z}))^{1-i_k} \right] dF_{\mathbf{Z}}(\mathbf{z}) \end{aligned}$$

- Recovers independence in the case \mathbf{Z} is degenerate.



Pairwise correlations

- Bivariate joint probabilities for any pairs:

$$P(I_k = 1, I_{k^*} = 1) = E_{\mathbf{Z}} [Q_k(\mathbf{Z}) Q_{k^*}(\mathbf{Z})]$$

- Covariance for any pairs:

$$\text{Cov}(I_k, I_{k^*}) = E_{\mathbf{Z}} [Q_k(\mathbf{Z}) Q_{k^*}(\mathbf{Z})] - E_{\mathbf{Z}} [Q_k(\mathbf{Z})] E_{\mathbf{Z}} [Q_{k^*}(\mathbf{Z})]$$

- Variance:

$$\text{Var}(I_k) = E_{\mathbf{Z}} [Q_k(\mathbf{Z})] (1 - E_{\mathbf{Z}} [Q_k(\mathbf{Z})])$$

- Pairwise correlation coefficient:

$$\rho(I_k, I_{k^*}) = \frac{\text{Cov}(I_k, I_{k^*})}{\sqrt{\text{Var}(I_k) \text{Var}(I_{k^*})}}$$



Ratio of relative risk

- Here we ask: “how much does one insurance (or credit) risk induce another insurance (or credit) risk to go on claim?”
- Ratio of relative risk:

$$\begin{aligned}\delta(I_k, I_{k^*}) &= \frac{P(I_k = 1 | I_{k^*} = 1)}{P(I_k = 0 | I_{k^*} = 1)} \\ &= \frac{E_{\mathbf{Z}}[Q_k(\mathbf{Z}) Q_{k^*}(\mathbf{Z})]}{E_{\mathbf{Z}}[(1 - Q_k(\mathbf{Z})) Q_{k^*}(\mathbf{Z})]}\end{aligned}$$

- An important measure of dependence for Bernoulli random variables!



Mixture models with covariates

- Covariates are often introduced into the model to capture the non-homogeneity in the portfolio and are used to understand how they influence the probability of a claim (or default).
- Denote the observed values of these r covariates by $\mathbf{x}_k = (x_{1k}, \dots, x_{rk})'$ and these covariates will enter into the model specification via:

$$P(I_k = 1 | \mathbf{Z}; \mathbf{x}_k) = p_{k|\mathbf{Z}}(1 | \mathbf{z}; \mathbf{x}_k) = Q(\mathbf{Z}; \mathbf{x}_k)$$

- Specified functional form for

$$Q(\mathbf{Z}; \mathbf{x}_k) = g(\mathbf{x}_k' \beta + \sigma' \mathbf{Z})$$

where $g : \mathbf{R} \rightarrow [0, 1]$ is some increasing function such as a distribution function (e.g. $g = \Phi$).



Specific model specifications

- Exponential-Gamma mixing distributions:

$Q(\mathbf{Z}; \mathbf{x}_k) = 1 - \exp(-\mathbf{x}'_i \beta - \sigma' \mathbf{Z})$ where \mathbf{Z} is a vector of independent Gamma distributed random variables.

- Logit-Normal mixing distributions:

$Q(\mathbf{Z}; \mathbf{x}_k) = 1 / (1 + \exp(-\mathbf{x}'_i \beta - \sigma' \mathbf{Z}))$ where \mathbf{Z} is a vector of independent standard Normally distributed random variables.

- Probit-Normal mixing distributions: $Q(\mathbf{Z}; \mathbf{x}_k) = \Phi(\mathbf{x}'_i \beta + \sigma' \mathbf{Z})$ where \mathbf{Z} is a vector of independent standard Normally distributed random variables.

- Beta-binomial model: $Q(Z) = Z$ where Z is a Beta(a, b) distributed random variable.

- covariates can be introduced: Prentice (1986, 1988).



Data description

- Data consists of policy exposure and claims experience:
 - portfolio of automobile insurance
 - period 1993-2001
 - insurance company in Singapore
- Records consisted of exposure and experience at the individual registered and insured vehicle level
 - details about driver and automobile risk characteristics
- Observable data:

$$\{I_{it}, e_{it}, \mathbf{x}_{it}, t = 1, \dots, T_i, i = 1, 2, \dots, m\}.$$



Claim frequency and exposure

Year	Number of claims	Policy exposed	Claims frequency
1993	840	12,157	6.9%
1994	1,739	15,389	11.3%
1995	869	8,074	10.8%
1996	736	7,556	9.7%
1997	1,760	16,216	10.9%
1998	2,455	23,691	10.4%
1999	3,630	36,647	9.9%
2000	3,770	45,806	8.2%
2001	3,349	44,910	7.5%
Total	19,148	210,446	9.1%



Selection of mixture models

- **Model 1:** *Logit-Normal mixing distribution*

$$Q(Z) = 1 / (1 + \exp(-\mu - \sigma Z))$$

where Z is standard Normal.

- **Model 2:** *Probit-Normal mixing distribution*

$$Q(Z) = \Phi(\mu + \sigma Z)$$

where Z is standard Normal.

- **Model 3:** *Beta mixing variable*
 - Q has Beta(a, b) distribution.
 - Appendix provides calculation details resulting from a Beta-Binomial model.



Maximum likelihood estimates

Mixture model	parms	estimates	standard errors	Neg log likelihood $-\log L(\theta; \mathbf{k})$	AIC	BIC
Logit-Normal	μ	-2.418	0.014 *	65,195.1	130,394	130,391
	σ	0.696	0.022 *			
Probit-Normal	μ	-1.398	0.007 *	65,154.8	130,314	130,310
	σ	0.342	0.012 *			
Beta-Binomial	a	13.6	0.125 *	65,292.4	130,589	130,586
	b	132.0	3.205 *			

* indicates significant at the 5% level.



Degrees of dependence

Mixture model	dependence measures	estimates	standard errors
Logit-Normal	q	0.0931	0.0007 *
	δ	0.1446	0.0032 *
	ρ	0.0367	0.0025 *
Probit-Normal	q	0.0934	0.0007 *
	δ	0.1477	0.0034 *
	ρ	0.0390	0.0027 *
Beta-Binomial	q	0.0936	0.0030 *
	δ	0.1108	0.0037 *
	ρ	0.0068	0.0002 *

* indicates significant at the 5% level.



Premium as a covariate

- We focus on the Logit-Normal and the Probit-Normal mixture models.
- The premium is introduced as a covariate through the μ parameter by specifying that

$$\mu = \beta_0 + \beta_1 \log(\text{Premium}/1000).$$

- For the Logit-Normal model, the intercept is -1.804 with s.e. of 0.014, and the coefficient of $\log(\text{Premium}/1000)$ is 0.528 with s.e. of 0.009.
- For the Probit-Normal model, the intercept is -1.080 with s.e. of 0.007. The coefficient of the $\log(\text{Premium}/1000)$ is 0.261 with s.e. of 0.004.



MLE with premium as covariate

Mixture model	parms	estimates	standard errors	Neg log likelihood $-\log L(\theta; \mathbf{k})$	AIC	BIC
Logit-Normal	β_0	-1.804	0.014 *	63,043.6	126,093	126,088
	β_1	0.528	0.009 *			
	σ	0.521	0.024 *			
Probit-Normal	β_0	-1.080	0.007 *	63,074.1	126,154	126,149
	β_1	0.261	0.004 *			
	σ	0.289	0.013 *			

* indicates significant at the 5% level.



Figure 1 - claim probability

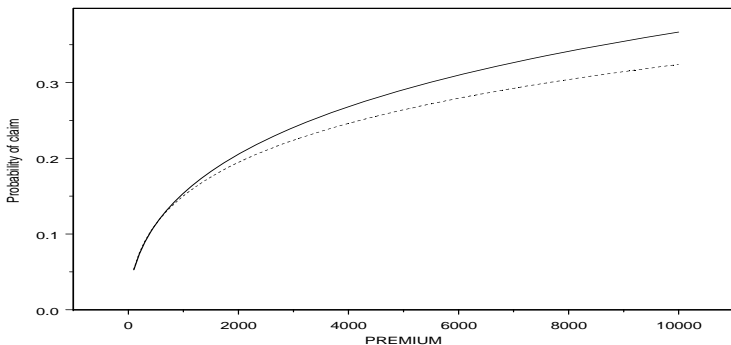


Figure: Claim probability - Logit-Normal vs. Probit-Normal



Figure 2 - relative risk

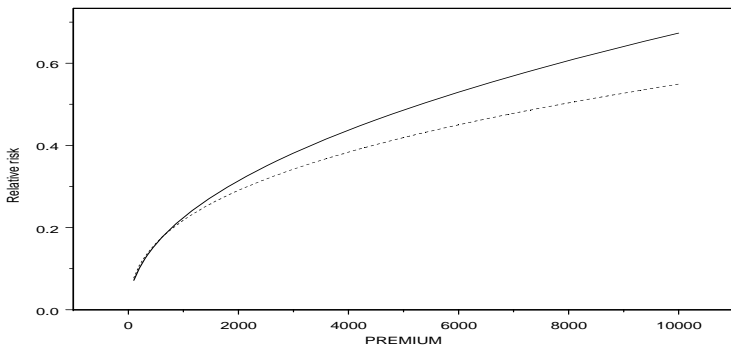


Figure: Relative risk - Logit-Normal vs. Probit-Normal



Figure 3 - correlation

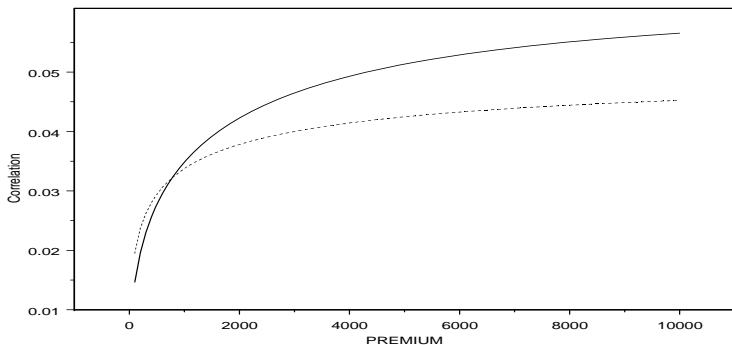


Figure: Correlation - Logit-Normal vs. Probit-Normal



Concluding remarks

- Mixture models reduce the dimensionality of the problem.
- Because the likelihood function requires integrating out the effect of the mixing variable, estimation routines can become cumbersome. However, SAS has procedure called NL MIXED that allow for this and that can accommodate covariates as well.
- A third advantage of the mixture model is some possible mathematical tractability.
- Finally, the unobservable variable in the Bernoulli mixture model has the advantage of providing a natural interpretation to the resulting model.



Acknowledgments

The author wishes to acknowledge

- Mitchell Wills for his research assistance on this work;
- Actuarial Education Research Fund (AERF) and the Casualty Actuarial Society (CAS) for funding this research project;
- Australian Research Council through the Discovery Grant DP0345036 and the UNSW Actuarial Foundation of the Institute of Actuaries of Australia for additional financial support.



Some references

Frey, R., McNeil, A.J. (2003) "Dependent Defaults in Models of Portfolio Credit Risk," *Journal of Risk* **6**: 59-92.

McNeil, A.J., Frey, R., Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.

Prentice, R.L. (1988) "Correlated Binary Regression with Covariates Specific to Each Binary Observation," *Biometrics* **44**: 1033-1048.

Valdez, E.A. (2006) "Empirical Investigation of Insurance Claim Dependencies", working paper, submitted for publication *Journal of Casualty Actuarial Science*.