Prediction of settlement delay in critical illness insurance

George Streftaris
Heriot-Watt University, Edinburgh

3rd EAJ, Lyon, September 2016

Work with:

Erengul Dodd (Ozkok), Southampton U

Research funded by UK Institute & Faculty of Actuaries projects:

2012-14 *Incorporating model and parameter uncertainty in rate graduation and pricing for CII*

2016-20 *Modelling, Measurement and Management of Longevity and Morbidity Risk*

Plan:

1. Critical Illness Insurance (CII) and data
2. Diagnosis-to-settlement delay distribution modelling
3. Model assessment and comparison
4. Prediction of settlement delay

Critical Illness: Policy description

- Fixed term policy, usually ceasing at age 65
- A fixed sum insured payable on the diagnosis of one of a specified list of critical illnesses
- Policies are often sold together with a term or an endowment insurance
- Benefit type:
  Full acceleration (FA): Death is included as a critical illness (88%)
  Stand alone (SA): Death is not included as a critical illness (12%)
- Covers:
  Cancer; Death; Heart attack; Stroke; Multiple Sclerosis; Total & permanent disability; Coronary artery bypass graft; Kidney failure; Major organ transplant; Other.

Data

CI data for 1999 – 2005 supplied to Heriot–Watt U by the CMI:

- Details of policies inforce at the start and end of each year

  $\rightarrow$ 18 000 000 policy-years of exposure
- Details of claims settled in 1999 – 2005

  $\rightarrow$ 19 000 claims

Covariates in the data:

Covariate	Number of levels
Age	Numerical
Sex	2 (Female = 0)
Smoker status	2 (NS = 0)
Policy duration	Numerical
Office	13
Benefit type	2 (FA = 0 & SA)
Benefit amount	Numerical
Policy type	2 (Single/Joint life = 0)
Settlement year	Numerical
Cause	10

Modelling diagnosis to settlement delay

- Diagnosis is the **insured event** and there is a **delay** between **diagnosis** and **settlement**
  – diagnosis date often not recorded (18%); need to model it
  – does delay also depend on risk factors?
- Observed data: Mean Delay = 185 days; SD Delay = 263 days

Modelling diagnosis to settlement delay

- Diagnosis is the **insured event** and there is a **delay** between **diagnosis** and **settlement**
  – diagnosis date often not recorded (18%); need to model it
  – does delay also depend on risk factors?
- Observed data: Mean Delay = 185 days;   SD Delay = 263 days
- Fit a delay distribution $F(d;x,z)$:

$$F(d;x,z) = \Pr[\text{claim diagnosed age } x\text{, covariates } z\text{, will be settled within } d \text{ days}]$$

- Also take into account uncertainty

Modelling diagnosis to settlement delay

- Diagnosis is the **insured event** and there is a **delay** between **diagnosis** and **settlement**

  – diagnosis date often not recorded (18%); need to model it
  – does delay also depend on risk factors?

- Observed data: Mean Delay = 185 days; SD Delay = 263 days

- Fit a delay distribution $F(d; x, z)$:

$$F(d; x, z) = \Pr[\text{claim diagnosed age } x \text{, covariates } z \text{, will be settled within } d \text{ days}]$$

- Also take into account uncertainty

Diagnosis-to-settlement Delay (D) distribution modelling

Include **risk factors** (covariates, z) in **GLM-type setting**:

M_1: $D_i \sim$ Generalised Beta2$(\alpha, \tau, \gamma, s_i)$

$$f_D(d_i) = \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \frac{\tau (d_i/s_i)^{\tau\gamma}}{d_i \left[1 + (d_i/s_i)^{\tau}\right]^{\alpha+\gamma}}$$

$$E(D_i) = \exp(\eta_i) = \exp\left(\beta_0 + \sum_{j=1}^{8} \beta_j z_{ij} + \beta_{9,k} + \beta_{10,l}\right)$$

with s_i given as function of $\eta_i, \alpha, \tau, \gamma$.

M_2: $D_i \sim$ Burr(α, τ, s_i)

As GB2 above, with $\gamma = 1$.

M_3: $D_i \sim$ Pareto(α, s_i)

As GB2 above, with $\tau = \gamma = 1$.

Diagnosis-to-settlement Delay (D) distribution modelling

Include **risk factors** (covariates, z) in **GLM-type setting**:

M_1: $D_i \sim$ Generalised Beta2$(\alpha, \tau, \gamma, s_i)$

$$f_D(d_i) = \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)} \frac{\tau (d_i/s_i)^{\tau\gamma}}{d_i \left[1 + (d_i/s_i)^{\tau}\right]^{\alpha+\gamma}}$$

$$E(D_i) = \exp(\eta_i) = \exp\left(\beta_0 + \sum_{j=1}^{8} \beta_j z_{ij} + \beta_{9,k} + \beta_{10,l}\right)$$

with s_i given as function of $\eta_i, \alpha, \tau, \gamma$.

M_2: $D_i \sim$ Burr(α, τ, s_i)

As GB2 above, with $\gamma = 1$.

M_3: $D_i \sim$ Pareto(α, s_i)

As GB2 above, with $\tau = \gamma = 1$.

Delay (D) distribution modelling (cont.)

M_4: $D_i \sim \mathrm{LN}(\mu_i, \sigma^2)$

$$E(D_i) = \exp(\eta_i + \sigma^2/2)$$

where

$$\eta_i = \mu_i = \beta_0 + \sum_{j=1}^{8} \beta_j z_{ij} + \beta_{9,k} + \beta_{10,l}$$

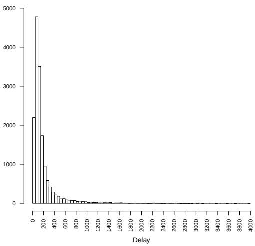
M_5: $D_i \sim$ Transformed (generalised) Gamma(α, τ, s_i)

$$f_D(d_i) = \frac{\tau (d_i/s_i)^{\alpha\tau} \exp(-d_i/s_i)^{\tau}}{d_i \Gamma(\alpha)} \quad , \quad E(D_i) = \exp(\eta_i)$$

where η_i as above and s_i given as function of η_i, α, τ.

Model fitting – without covariates

Fit the 5 **null** models under a Bayesian framework using Markov chain Monte Carlo



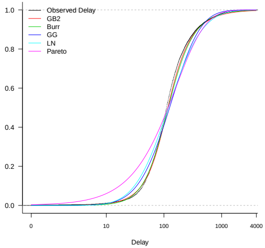


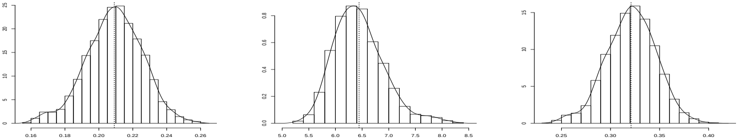
Figure: Left: histogram of the observed delay (in days). Right: CDF of observed delay (in days, on log scale) and fitted distributions.

- GB2 and Burr similar fit - especially at tails
- Pareto least successful

Model fitting – with covariates

Include risk factors (covariates) z: age, sex, smoking, cause etc ...

Posterior estimates of GB2 parameters



(a) α (b) τ (c) γ

Posterior estimates here support GB2 as opposed to:
Pareto ($\tau = \gamma = 1$), or
Burr ($\gamma = 1$).

Covariate coefficient estimates

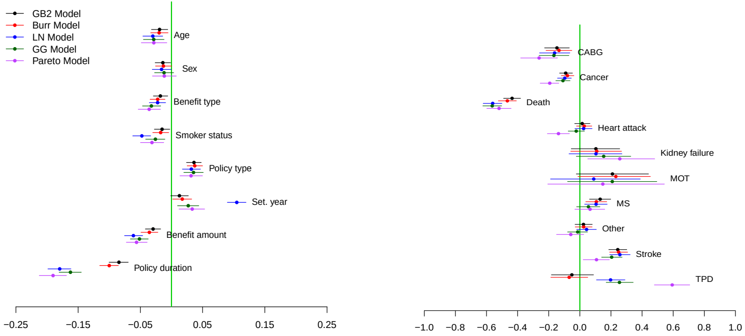


Figure: Posterior means (dots) and 95% credible intervals (bars) of β's.

- Similar estimates, especially between GB2 and Burr
- GB2 more efficient with missing values (smaller sd – not shown here)
- Some covariates more affected by tail structure of distn (eg settlement year, pol duration)

Model assessment & comparison

Important for prediction, but not straightforward - especially with missing data.

DIC (with missing values) much criticised.

	GB2	Burr	GG	Log-normal	Pareto
DIC_4	**230,952**	231,251	233,262	237,798	237,665
DIC_5	**231,677**	232,002	233,835	238,765	238,297
DIC_8	**191,037**	191,315	193,065	194,796	196,060

Instead consider:

- Latent likelihood ratio (LLR) tests
- Bayesian latent 'residuals' (BLR) based on cdf

Model assessment & comparison (cont)

- Both methods originally developed for epidemics (eg Streftaris & Gibson, 2012)
- Result in post distns of *p*-values
- Evidence against a model coming from post distn concentrated close to zero.

Latent likelihood ratio tests:
When comparing Burr to GB2 model, all $\pi_\Lambda^{(t)} \approx 0$
$\longrightarrow$ **overwhelming evidence in favour of GB2**

Bayesian latent 'residuals':
Under Burr 100% of π_Q values smaller than 5×10^{-9}
$\longrightarrow$ **again evidence against Burr**

Model assessment & comparison (cont)

- Both methods originally developed for epidemics (eg Streftaris & Gibson, 2012)
- Result in post distns of *p*-values
- Evidence against a model coming from post distn concentrated close to zero.

Latent likelihood ratio tests:

When comparing Burr to GB2 model, all $\pi_\Lambda^{(t)} \approx 0$
$\longrightarrow$ **overwhelming evidence in favour of GB2**

Bayesian latent 'residuals':

Under Burr 100% of $\boldsymbol{\pi}_Q$ values smaller than 5×10^{-9}
$\longrightarrow$ **again evidence against Burr**

Variable selection with GB2

$2^{10} = 1024$ possible models – Use Gibbs variable selection

Model probabilities:

Model		$\hat{f}(m\|\boldsymbol{D})$	$PO(m_{977}/.)$
m_{977}	$z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1996	1.00
m_{981}	$z_3 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1843	1.08
m_{978}	$z_1 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1503	1.33
m_{982}	$z_1 + z_3 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1007	1.98
m_{979}	$z_2 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.0449	4.44

(1) age; (2) sex; (3) benefit (FA/SA); (4) smoking; (5) policy type; (6) year; (7) amount; (8) duration; (9) office; (10) cause

Small difference in PO and probabilities among first 4 models
$\longrightarrow$ Use model averaging for prediction

Variable selection with GB2

$2^{10} = 1024$ possible models – Use Gibbs variable selection

Model probabilities:

Model		$\hat{f}(m \mid \boldsymbol{D})$	$PO(m_{977}/.)$
m_{977}	$z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1996	1.00
m_{981}	$z_3 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1843	1.08
m_{978}	$z_1 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1503	1.33
m_{982}	$z_1 + z_3 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.1007	1.98
m_{979}	$z_2 + z_5 + z_7 + z_8 + z_9 + z_{10}$	0.0449	4.44

(1) age; (2) sex; (3) benefit (FA/SA); (4) smoking; (5) policy type; (6) year; (7) amount; (8) duration; (9) office; (10) cause

Small difference in PO and probabilities among first 4 models
$\longrightarrow$ Use model averaging for prediction

Prediction of Claim Delay

Consider following **missing** delays and post **predictions** under the 'best' (m_{977}) and average model:

	1	2	3	4
Type	JL	JL	JL	JL
£	50 000	50 000	**5 000**	**115 000**
Durn (yrs)	> 5	$<$**1**	> 5	> 5
Office	2	2	2	2
Cause	Death	Death	Death	Death
Type	FA	FA	FA	FA
		Prediction		
m_{977}	221.2 (203.6, 240.2)	271.6 (251.4, 292.7)	234.6 (216.8, 254.2)	214.1 (196.2, 232.7)
Average model	220.3 (202.9, 236.2)	270.4 (249.9, 289.1)	233.5 (215.0, 250.4)	213.3 (196.2, 228.9)

Boldface: changes from reference case (1&6)

Prediction of Claim Delay (cont.)

Missing delays and post **predictions** under the 'best' (m_{977}) and average model:

	5	6	7	8
Type	JL	**SL**	SL	SL
£	50 000	50 000	50 000	50 000
Durn (yrs)	> 5	> 5	> 5	> 5
Office	2	2	**11**	2
Cause	**MS**	Death	Death	**Cancer**
Benefit	FA	FA	FA	SA
		Prediction		
m_{977}	403.4 (368.4, 440.6)	244.6 (225.5, 265.1)	152.6 (141.1, 163.6)	323.4 (298.9, 351.0)
Average model	402.4 (366.8, 439.3)	244.1 (224.8, 262.8)	152.7 (141.2, 163.6)	315.8 (292.2, 339.4)

Boldface: changes from reference case (1&6)

Summary

- Estimation & prediction of diagnosis-to-settlement delay **important in CII**
- Bayesian analysis **accounts for non-recorded diagnosis** dates
- Previous work has shown that estimates of delay are model-sensitive
- 4-parameter **GB2 distn** most suitable (as shown by using various methods)
- Variable selection leads to **model-averaged prediction** for non-recorded delays
- Results here feed in work on CII claim rates
- Future/continuing work to involve more recent data

More details in:

Ozkok, E., Streftaris, G., Waters, H.R., and Wilkie, A.D. (2012)
Bayesian modelling of the time delay between diagnosis and settlement for Critical Illness Insurance using a Burr generalised-linear-type model. *Insurance: Mathematics & Economics*, 50, 266–279.

Streftaris G. and Gibson, G.J. (2012)
Non-exponential tolerance to infection in epidemic systems modelling, inference and assessment, *Biostatistics*, 13, 580593

Ozkok, E., Streftaris, G., Waters, H.R., and Wilkie, A.D. (2014)
Modelling critical illness claim diagnosis rates I: Methodology. *Scandinavian Actuarial Journal*, 2014:5, 439–457.

Dodd, E., Streftaris, G., Waters, H.R. and Stott, A.D. (2015)
The effect of model uncertainty on the pricing of critical illness insurance. *Annals of Actuarial Science*, 9, 108–133.

Dodd, E. and Streftaris, G. (2016)
Prediction of settlement delay in critical illness insurance claims using GB2 distribution. To appear in *Journal of the Royal Statistical Society C*.

Model fitting – with covariates

Include risk factors (covariates) z: age, sex, smoking, cause etc ...

GL-type model linked to the mean of the GB2 through

$$\log\left(E(D_i)\right) = \eta_i = \beta_0 + \sum_{j=1}^{8} \beta_j z_{ij} + \beta_{9\,O_i} + \beta_{10\,C_i}$$

Use MCMC to draw samples from posterior

$$f(\alpha, \tau, \gamma, \boldsymbol{\beta} | \boldsymbol{D}) \propto f(\boldsymbol{D} | \alpha, \tau, \gamma, \boldsymbol{\beta}) f(\alpha) f(\tau) f(\gamma) f(\boldsymbol{\beta})$$

with (mainly vague) priors:

$$\alpha \sim \text{Gamma}(1, 0.01) I(1/\tau, \infty)$$
$$\tau \sim \text{Gamma}(1, 0.01)$$
$$\gamma \sim \text{Gamma}(1, 0.01)$$
$$\beta_j \sim \text{N}(0, 10^4), j = 1, \ldots, 8$$
$$\beta_{9\,O_i} \sim \text{N}(0, 10^4), O_i = 2, \ldots, 13$$
$$\beta_{10\,C_i} \sim \text{N}(0, 10^4), C_i = 2, \ldots, 10.$$

Latent likelihood ratio (LLR) tests

Originally developed for epidemics (eg Streftaris & Gibson, 2012)

- Fit model $\mathcal{M}_1$ (eg Burr) under Bayesian estimation
- For alternative model, $\mathcal{M}_2$ (GB2), compute ML value
- Calculate LLR at iteration t of MCMC

$$\Lambda^{(t)} = \frac{L_1\left(\alpha^{(t)}, \tau^{(t)}, \boldsymbol{\beta}^{(t)}; \boldsymbol{D}\right)}{L_2\left(\dot{\alpha}, \dot{\tau}, \dot{\gamma}, \dot{\boldsymbol{\beta}}; \boldsymbol{D}\right)}$$

  where $\boldsymbol{\alpha}^{(t)}, \tau^{(t)}, \boldsymbol{\beta}^{(t)}$ are MCMC posterior estimates at iteration (t) and dotted values are MLEs

- Evidence against $\mathcal{M}_1$ can be provided by tail probability
  $\pi_\Lambda^{(t)} = P\left(\Lambda \leq \Lambda^{(t)} | \mathbf{d}\right) \approx P(\chi^2_{df} \geq -2\log\Lambda^{(t)})$
  where df is the number of estimated parameters in $\mathcal{M}_2$.

When comparing Burr to GB2 model, all $\pi_\Lambda^{(t)} \approx 0$
$\longrightarrow$ **overwhelming evidence against Burr**.

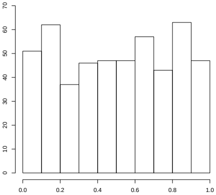
Posterior distns of p-values

Related to posterior predictive checking.

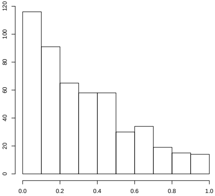
- $f(D_j|\boldsymbol{\theta}^{(t)})$ is the sampling distribution (eg GB2) for delay j at MCMC iteration t, $j = 1, \ldots, k$ and $t = 1, \ldots, N$
- Compute cdf value $q_j^{(t)} = P(D_j \leq D_j^{\text{obs}}|\boldsymbol{\theta}, \boldsymbol{D})$
- Under hypothesis that model fits data adequately: $\boldsymbol{q}^{(t)} = q_1^{(t)}, \ldots, q_k^{(t)} \sim U(0,1)$
- Obtain p-value $\pi_Q^{(t)}$ for compliance with $U(0,1)$ at each MCMC iteration (e.g. KS g-o-f test)
- $\boldsymbol{\pi}_Q = \pi_Q^{(1)}, \ldots, \pi_Q^{(N)}$ is sample from post distn of $\boldsymbol{\pi}_Q$ and is used for evidence against fitted model

Posterior distns of p-values (cont)

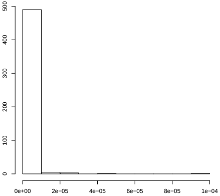
Apply to simulated data ($k = 500$)



(a) GB2

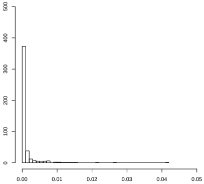


(b) Burr

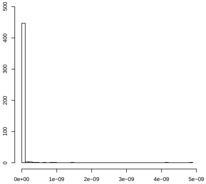


(c) Pareto

With observed data ($k = 19,127$)



(d) GB2



(e) Burr

Under Burr 100% of π_Q values smaller than 5×10^{-9}