Online Appendix for 'Joint model prediction and application to individual-level loss reserving', ASTIN Bulletin

Online Appendices

A: Alternative Approaches to Estimating Joint Model

In this section, we investigate two alternative estimating strategies for the joint model. We show that significant bias will be induced if the model is not estimated appropriately.

The first strategy is independence estimation. Specifically, setting $\alpha = 0$ in the survival submodel, we estimate the longitudinal and survival submodel separately. The second strategy is a two-stage method. The first stage estimates the longitudinal submodel, and the second stage estimates the survival submodel holding parameter estimates from the first stage fixed.

In the simulation study, we generate data from the joint model described in Section 5.1. Parameters are estimated using both independence and two-stage methods. Estimation results based on sample size N=1000 and S=150 replications are reported in Table 1 and Table 2, respectively, for the longitudinal and survival submodels. We show in the table the average bias (Bias), the average standard error (SE) of the estimates, and the standard deviation of the average bias calculated as SD/$\sqrt{(S)}$. For comparison, we reproduce the estimates for the joint model from Table 1 in the paper.

It is critical to note that both estimation strategies induce substantial bias into parameter estimates. For the independence method, we emphasize that it is different from the usual multivariate regression where ignoring the association among multiple outcomes pays no price in terms of consistency, but only hampers the efficiency. The bias in the longitudinal submodel is due to the sample selection under independence assumption and the bias in the survival submodel is due to the omitted variable. For the two-stage estimation, the selection bias in the longitudinal submodel is carried over to the survival submodel. Therefore, model parameters cannot be consistently estimated although the association between the two processes are taken into account.

Table 1: Estimation results for the longitudinal submodel.

N=1000, S=150	JM			Independence and Two-stage		
Parameter	Bias	SD/$\sqrt{(S)}$	SE	Bias	SD/$\sqrt{(S)}$	SE
β_{10}=1.0	0.001	0.005	0.056	-0.035	0.005	0.056
β_{11}=0.3	0.002	0.001	0.010	0.008	0.001	0.010
β_{21}=0.2	-0.001	0.003	0.039	0.001	0.003	0.040
β_{22}=0.4	-0.002	0.003	0.042	-0.001	0.003	0.042
ν=0.09	-0.001	0.001	0.015	0.002	0.001	0.017
σ=1.5	0.001	0.004	0.043	0.092	0.006	0.072

Table 2: Estimation results for the survival submodel.

N=1000, S=150	JM			Two-Stage			Independence		
Parameter	Bias	SD/$\sqrt{(S)}$	SE	Bias	SD/$\sqrt{(S)}$	SE	Bias	SD/$\sqrt{(S)}$	SE
γ_1=0.5	-0.004	0.007	0.085	-0.004	0.007	0.086	-0.026	0.007	0.089
γ_2=0.3	-0.001	0.007	0.079	-0.004	0.007	0.080	-0.084	0.007	0.081
log(λ)=-1.139	-0.021	0.012	0.148	-0.075	0.012	0.140	-0.420	0.009	0.116
α=-0.25	0.011	0.005	0.066	0.020	0.005	0.063	-	-	-

B: Details for Marked Poisson Process for RBNS

Under the Marked Poisson Process framework, the likelihood for the full development process of a claim is given by (Jin, 2014):

$$L = f_V \times f_{U|v} \times f_{W|v,u} = f_V \times f_{U|v} \times f_{S|v,u} \times f_{E|v,u,s} \times f_{P|v,u,s,e}, \tag{1}$$

where V and U represent the claim occurrence times and reporting delay respectively. However, with the focus on RBNS reserve prediction, we are interested in the claim development process W given by

$$f_{W|v,u} = f_{S|v,u} \times f_{E|v,u,s} \times f_{P|v,u,s,e}. \tag{2}$$

Where S denotes the transaction occurrence times, E denotes the type of transaction, and P denotes the payment amount of the transaction. The transaction occurrence times S are modeled by a discrete survival model with piecewise constant hazard rates. Following Jin (2014) and Antonio and Plat (2014), the first transactions are modeled with a hazard rate $g(s)$, and the later transactions are modeled with a different hazard rate $h(s)$. Let $[0, a_R]$ and $[0, b_L]$ be the interval for first and later transactions. Then we have:

$$g(s) = \sum_{r=1}^{R} g_r 1\{a_{r-1} < s \leq a_r\} \tag{3}$$

$$h(s) = \sum_{l=1}^{L} h_l 1\{b_{l-1} < s \le b_l\} \tag{4}$$

With cumulative hazard rates given by:

$$G(s) = \int_0^s g(t)dt \tag{5}$$

$$H(s) = \int_0^s h(t)dt \tag{6}$$

Then the cumulative density functions of transaction occurrence times are given by:

$$\Pr(S_1 \le s) = 1 - \exp(-G(s)) \tag{7}$$

$$\Pr(S_k \le s) = 1 - \exp(-H(s)), \quad k > 1 \tag{8}$$

Let $a_R = N_1$ be regarded as the maximum waiting time to the first transaction, and $b_L = N_2$ is regarded as the maximum settlement delay. Then under these additional assumptions, the probability that the first transaction occurs at time $r, r = 1, 2, ..., N_1$ is

$$\Pr(S_1 = r | S_1 \le N_1) = \frac{\exp\{-G(r-1)\} - \exp\{-G(r)\}}{1 - \exp\{-G(N_1)\}} \tag{9}$$

And given the occurrence time of the first transaction, $S_{k-1} = c$, the probability that transaction k occurs at time $r, r = c+1, c+2, ..., N_2$ is

$$\Pr(S_k = r | S_{k-1} = c, S_k \le N_2) = \frac{\exp\{-H(r-1)\} - \exp\{-H(r)\}}{\exp\{-H(c)\} - \exp\{-H(N_2)\}} \tag{10}$$

For the Wisconsin LGPIF training dataset, the maximum waiting time for the first transaction is 17 months, and the maximum settlement delay is 27 months. It is assumed that there is at most one transaction in each month, and the transactions can only occur at the end of a month. As noted in Jin (2014), this discrete setup is consistent with the fact that many insurers aggregate transactions on a monthly basis by the end of each month. Therefore, the piecewise-constant hazard rates is defined to have jumps every month, i.e. $a_1 = 0, a_2 = 1, \cdots, a_{17} = 17$ and $b_1 = 0, b_2 = 1, \cdots, b_{27} = 27$.

Furthermore, for the type of transactions E, we consider two types for claim i at time $S = s$; a payment transaction that leads to settlement ($e_{is} = 1$) and an intermediate payment transaction ($e_{is} = 0$). With an intermediate transaction, the claim development process continues. Given a transaction at time s, the transaction type is determined by a logit model that accommodates heterogeneity by incorporating random effects a_i.The probabilities also depend on the time of the transaction and covariates x_{is} given by:

$$\Pr(e_{is} = 1|a_i) = \pi(x'_{is}\beta + a_i) = \frac{1}{1 + \exp(-(x'_{is}\beta + a_i))}. \tag{11}$$

To model the incremental payments P, a Generalized Linear Mixed-Effects Model is specified.

C: Joint Model Algorithm to Construct the Training and Validation Data in the Simulation Study

Algorithm 1 Data-generating process for JM.

Input: Parameters $\{\beta_{10}, \beta_{11}, \beta_{21}, \beta_{22}, \nu, \sigma\}$ from the payments submodel, and $\{\gamma_1, \gamma_2, \lambda, \alpha\}$ from the settlement submodel.

Output: Training datasets $D_T^P = \{(y_{it}, t, x_{i1}, x_{i2}); 0 \le t \le t_i, i = 1, ..., N\}$, and $D_T^S = \{(t_i, \delta_i, x_{i1}, x_{i2}); i = 1, ..., N\}$; Validation dataset for open claims $D_V = \left\{y_{it}; c_i < t \le t_i^*, i = 1, ..., m\right\}$.

for Claim $i = 1, ..., N$ **do**
  Generate $x_{i1} \sim$ Bernouli(0.5), $x_{i2} \sim$ Normal$(1, 0.25)$;
  Generate $\mathbf{b}_i = b_{i0} \sim \mathcal{N}(\mathbf{0}, \nu)$;
  for Payment time $t = 0, ..., 9$ **do**
    $y_{it} \sim \text{Gamma}\left(\frac{\exp(\eta_{it})}{\sigma}, \sigma\right)$; $\eta_{it} = \beta_{10} + t\beta_{11} + x_{i1}\beta_{12} + x_{i2}\beta_{13} + b_{i0}$;
  end for
  return $\{y_{it}; t = 0, ..., 9\}$;
  Generate $S_i(t) = U \sim$ Uniform$(0, 1)$;
  Calculate $t_i^* = H_i^{-1}(-\log(U))$; where $H_i(t) = \int_0^t \lambda \exp\{\gamma_1 x_{i1} + \gamma_2 x_{i2} + \alpha\eta_{is}\} ds$;
  Generate accident year $AY_i \in [(1, \ldots, 10) - 1]$;
  Generate $c_i = 9 - AY_i +$ Uniform$(0, 0.5)$;
  $D_{Ti}^P = \{(y_{it}, t, x_{i1}, x_{i2}); 0 \le t \le t_i\}$; where $t_i = \min(t_i^*, c_i)$;
  $D_{Ti}^S = \{t_i, \delta_i, x_{i1}, x_{i2}\}$; where $\delta_i = I(t_i^* < c_i)$;
  $D_{Vi} = \left\{y_{it}; c_i < t \le t_i^*, \delta_i = 0\right\}$;
  return $D_T^P = \left\{D_{Ti}^P; i = 1, ..., N\right\}$; $D_T^S = \left\{D_{Ti}^S; i = 1, ..., N\right\}$; and $D_V = \{D_{Vi}; i = 1, ..., m\}$; where $m = \sum I(\delta_i = 0)$;
end for