Simulation report – prediction of survival probabilities

Ming Yang

September 10, 2015

1 Objective

To compare the predictive capabilities of survival probabilities between QRJM and JM using linear mixed model (LMJM) for data from different distribution features.

2 Simulation procedure

1. Define different simulation scenarios in terms of the distribution of random error and simulate 100 data sets for each scenario (see below for the specification of scenarios). Each simulated data set has 600 subjects, 500 out of which will be used to fit the model for inference purpose and the rest 100 will be used to make predictions and validation.
2. Fit the data using QRJM and LMJM respectively and save the posterior samples of the model parameters.
3. Validation data preparation: choose a time t so that all the patients selected to for prediction will only have longitudinal measurements up to this time t.
4. Make predictions of subject-specific random effects: use saved posterior samples in step 2 and longitudinal measurements from step 3 to predict subject-specific random effects for every subject in the validation samples.
5. Calculate the predictions of survival probabilities for all the subjects in validation data for some time $u = t + \Delta t$ ($\Delta t > 0$).
6. Summarize the result: make Bland-Altman plots and calculate the MSE and bias for our predictions versus the gold standard, which is calculated from the true simulated values (i.e. the random effects and the parameters).

3 Simulation scenarios and results

$$\begin{cases} Y_{it} = \boldsymbol{X}_{it}^{\top}\boldsymbol{\beta} + \boldsymbol{H}_{it}^{\top}\boldsymbol{\delta} + \boldsymbol{Z}_{it}^{\top}\boldsymbol{u}_i + \varepsilon_{it}, \varepsilon_{it} \sim ALD(0, \sigma, \tau) \\ h(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i; \boldsymbol{\gamma}, \alpha_1, \alpha_2) = h_0(T_i)\exp(\boldsymbol{W}_i^{\top}\boldsymbol{\gamma} + \alpha(\boldsymbol{H}_{iT_i}^{\top}\boldsymbol{\delta} + \boldsymbol{Z}_{iT_i}^{\top}\boldsymbol{u}_i)) \end{cases} \tag{1}$$

There are three scenarios in the simulation study:

1. Scenarios One: data are generated using Model (1). Choose the $\tau = \mathbf{0.25}$ for the ALD distribution.

2. Scenarios Two: data are generated using Model (1). Choose $\tau =$**0.5** for the ALD distribution.
3. Scenario Three: data are generated using Model (1), but the random error follows standard normal distribution instead of ALD.

3.1 Inference results

Table 1: Inference result for Scenario One

	QRJM ($\tau = 0.25$), true model			QRJM ($\tau = 0.5$)			LMJM		
	bias	se	MSE	bias	se	MSE	bias	se	MSE
alpha	0.008	0.110	0.012	-0.003	0.131	0.017	-0.012	0.138	0.019
beta[1]	-0.001	0.101	0.010	1.697	0.156	2.904	2.683	0.174	7.227
beta[2]	-0.005	0.096	0.009	0.002	0.132	0.017	0.014	0.150	0.023
c	-0.007	0.090	0.008	-0.004	0.091	0.008	-0.007	0.090	0.008
delta[1]	0.002	0.085	0.007	0.019	0.113	0.013	0.029	0.129	0.018
delta[2]	0.009	0.092	0.009	0.018	0.110	0.012	0.038	0.128	0.018
gamma[1]	0.007	0.083	0.007	0.012	0.086	0.007	0.007	0.085	0.007
gamma[2]	-0.001	0.087	0.008	0.005	0.090	0.008	0.000	0.089	0.008
sigma	-0.001	0.034	0.001	-0.319	0.025	0.103	-	-	-

Table 2: Inference result for Scenario Two

	QRJM ($\tau = 0.5$), true model			LMJM		
	bias	se	MSE	bias	se	MSE
alpha	0.013	0.094	0.009	0.013	0.106	0.011
beta[1]	-0.007	0.089	0.008	-0.009	0.112	0.013
beta[2]	0.011	0.080	0.007	0.013	0.103	0.011
c	0.002	0.084	0.007	-0.001	0.086	0.007
delta[1]	-0.009	0.075	0.006	-0.006	0.092	0.009
delta[2]	0.002	0.082	0.007	0.006	0.097	0.009
gamma[1]	0.009	0.090	0.008	0.009	0.090	0.008
gamma[2]	0.001	0.086	0.007	0.002	0.087	0.008
sigma	0.003	0.037	0.001	-	-	-

Table 3: Inference result for Scenario Three

	QRJM ($\tau = 0.5$)			LMJM, true model		
	bias	se	MSE	bias	se	MSE
alpha	-0.012	0.075	0.006	0.004	0.076	0.006
beta[1]	0.013	0.050	0.003	0.003	0.046	0.002
beta[2]	0.001	0.045	0.002	0.002	0.043	0.002
c	0.001	0.082	0.007	0.005	0.081	0.007
delta[1]	0.007	0.045	0.002	0.001	0.041	0.002
delta[2]	0.008	0.058	0.003	0.000	0.055	0.003
gamma[1]	-0.003	0.081	0.007	-0.007	0.080	0.006
gamma[2]	-0.002	0.085	0.007	-0.006	0.084	0.007
sigma	-	-	-	0.001	0.029	0.001

3.2 BA plots of predictions

Scenario One

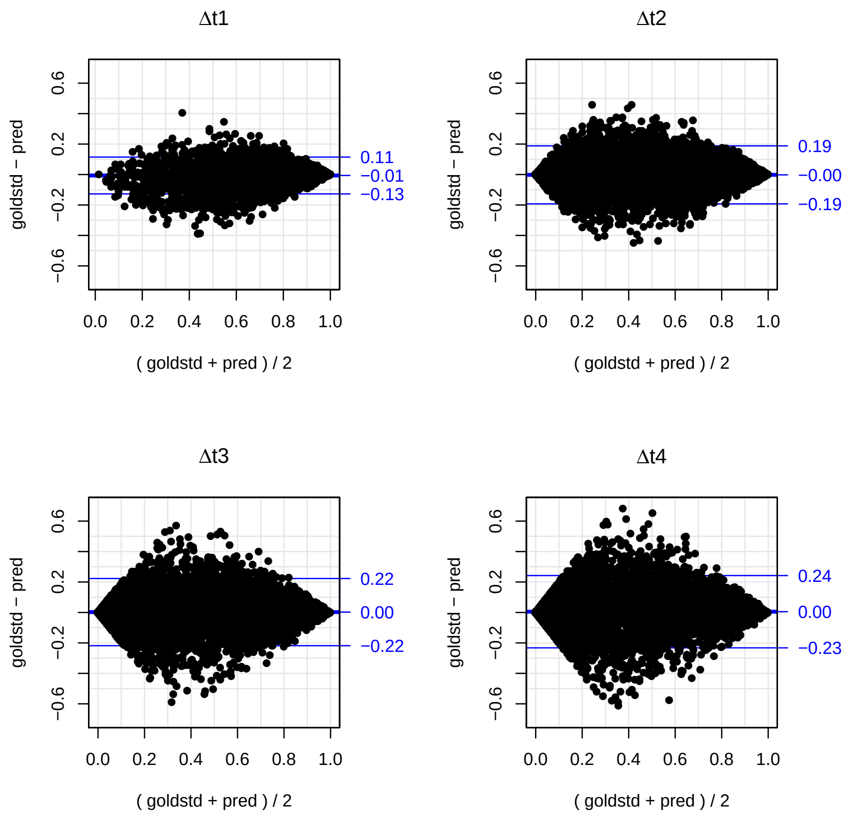

Figure 1: BA plot: data fitted using QRJM model with $\tau = 0.25$

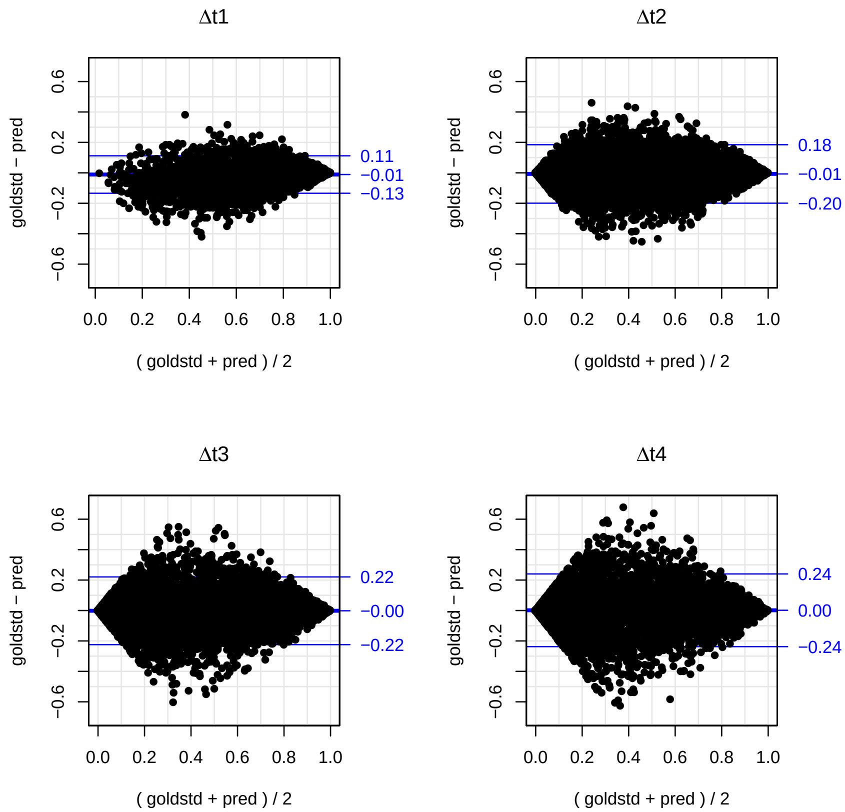

Figure 2: BA plot: data fitted using QRJM model with $\tau = 0.50$

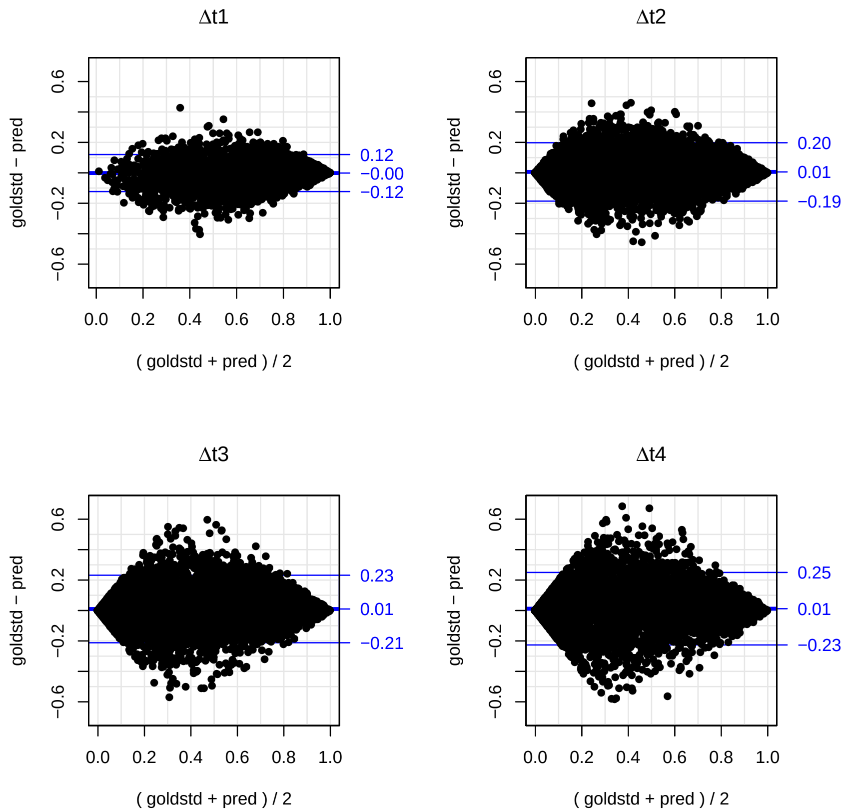

Figure 3: BA plot: data fitted using LMJM

Scenario Two

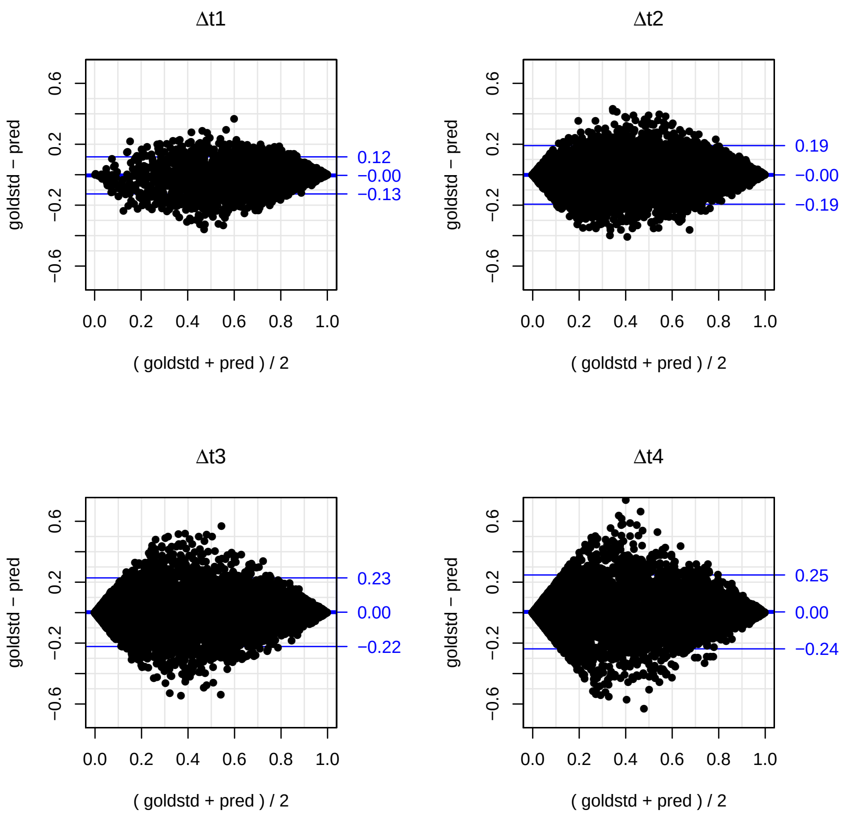

Figure 4: BA plot: data fitted with JM using QR model with $\tau = 0.5$

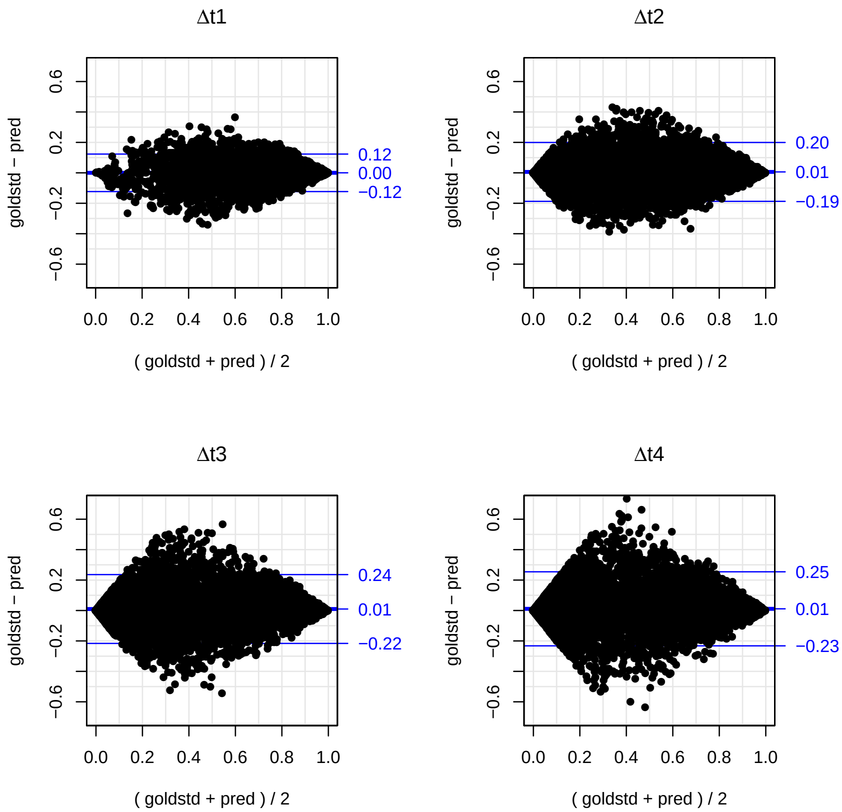

Figure 5: BA plot: data fitted using LMJM

Scenario Three

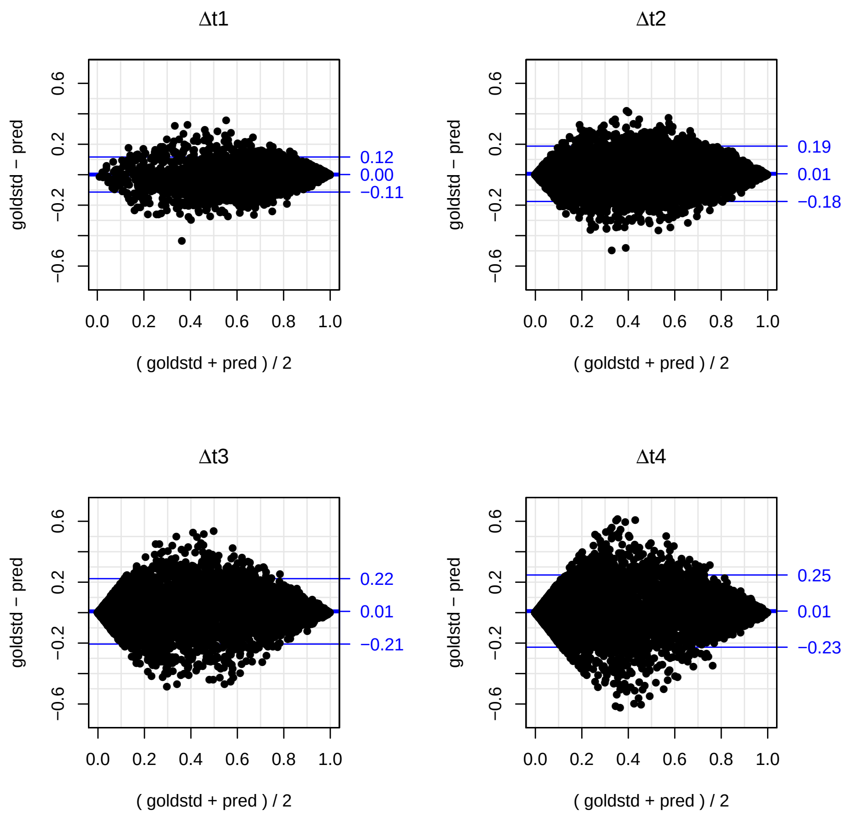

Figure 6: BA plot: data fitted using LMJM

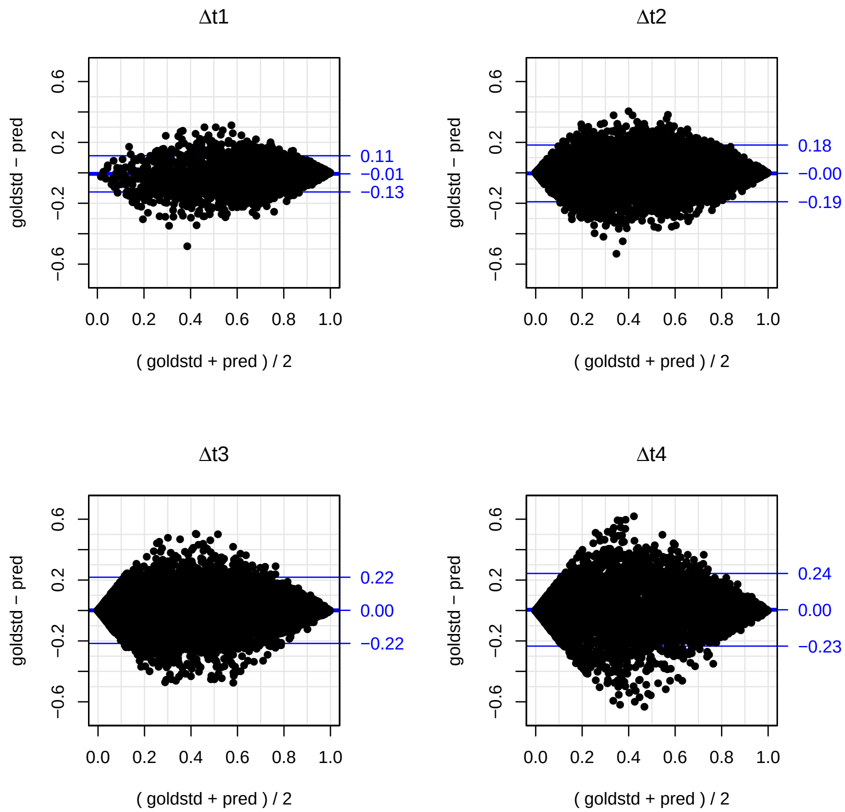

Figure 7: BA plot: data fitted with JM using QR model with $\tau = 0.5$

3.3 Summary table – predictions

Table 4: MSE and bias of the predictions of survival probabilities from two models ($\Delta t_1 < \Delta t_2 < \Delta t_3 < \Delta t_4$)

Scenario		QRJM ($\tau = 0.25$)		QRJM ($\tau = 0.5$)		LMJM	
		MSE	Bias	MSE	Bias	MSE	Bias
1	Δt_1	0.004	0.006	0.004	0.011	0.004	0.001
	Δt_2	0.009	0.002	0.009	0.007	0.009	-0.006
	Δt_3	0.012	-0.002	0.012	0.002	0.012	-0.010
	Δt_4	0.014	-0.005	0.014	-0.001	0.014	-0.012
2	Δt_1			0.004	0.004	0.004	0.000
	Δt_2	-	-	0.009	0.001	0.009	-0.006
	Δt_3	-	-	0.013	-0.003	0.013	-0.010
	Δt_4	-	-	0.015	-0.005	0.015	-0.011
3	Δt_1			0.004	0.007	0.003	-0.001
	Δt_2	-	-	0.009	0.002	0.004	-0.006
	Δt_3	-	-	0.012	-0.012	-0.001	-0.008
	Δt_4	-	-	0.014	-0.012	-0.005	-0.010

3.4 Comments

1. In both model settings, the accuracy (in terms of MSE and the BA plots) of predictions decreases as Δt increases, which makes sense as it's more difficult to accurately predict the survival for longer time in the future since there are more variabilities and uncertainty.

2. QRJM and LMJM perform closely in predicting the survival probabilities in all three scenarios although the inference results from the true model is better than others. I have no formal interpretation yet but thinking that probably it is because in the prediction part we only use parameters α, δ and γ to calculate the survival probabilities (But we used all the model parameters in predicting random effects), whose estimation from both model settings are pretty accurate and close to truth.