Dynamic Empirical Bayes Models and Their Applications to Finance and Insurance

Tze Leung Lai, Stanford University

Joint work with Yong Su, Kevin Sun

December 19, 2011

Outline

- Empirical Bayes (EB) methods and credibility models in insurance
- Evolutionary credibility and dynamic EB models
- Linear dynamic EB via linear mixed models (LMM)
  - Application to baseball batting averages
- Generalized linear mixed models (GLMM) and dynamic EB
  - Worker's compensation insurance
  - Baseball batting averages revisited
  - Default modeling of corporate bonds
- Conclusion

Empirical Bayes Methodology

- Empirical Bayes (EB) methods (Robbins, Stein)
  - EB replaces the hyperparameters of a Bayes procedure by maximum likelihood, method of moments or other estimates from the data.
  - These methods allow one to estimate statistical quantities (probabilities, functions of parameters, etc.) of an individual by combining information from the individual and other subjects in an empirical study.
- Hyperparameter estimation
  - Nonparametric empirical Bayes (Robbins: Poisson rates)
  - Parametric empirical Bayes (Stein, James & Stein, Efron & Morris: normal means)

Insurance Rate-Making: Credibility Models

- Standard credibility models (Bühlmann & Gisler, 2005) are essentially linear empirical Bayes.
- Suppose there are I risk classes and let Y_{ij} denote the j^{th} claim of the i^{th} class. Assume that (Y_{ij}, θ_i) are independent with $\mathrm{E}[Y_{ij}|\theta_i] = \theta_i$ and $\mathrm{Var}[Y_{ij}|\theta_i] = \sigma_i^2$, $(1 \le j \le n_i,\ 1 \le i \le I)$.
- Assuming a normal prior $N(\mu, \tau^2)$ for θ_i, the Bayes estimate of θ_i (that minimizes the Bayes risk) is

$$\mathrm{E}[\theta_i | Y_{i1}, \cdots, Y_{i,n_i}] = \alpha_i \bar{Y}_i + (1 - \alpha_i)\mu,$$

where $\alpha_i = \tau^2/(\tau^2 + \sigma_i^2/n_i)$ and $\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$.

Insurance Rate-Making: Credibility Models

- Since
$$\mathrm{E}[Y_{ij}] = \mathrm{E}[\mathrm{E}[Y_{ij}|\theta_i]] = \mathrm{E}[\theta_i] = \mu,$$
$$\mathrm{Var}[Y_{ij}] = \mathrm{Var}[\theta_i] + \mathrm{E}[\mathrm{Var}[Y_{ij}|\theta_i]] = \tau^2 + \sigma_i^2,$$
we can estimate μ, σ_i^2 and τ^2 by the method of moments:
$$\hat{\mu} = (\textstyle\sum_{i=1}^{I} \sum_{j=1}^{n_i} Y_{ij}) / \sum_{i=1}^{I} n_i,$$
$$\hat{\sigma}_i^2 = \textstyle\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 / (n_i - 1),$$
$$\hat{\tau}^2 = \textstyle\sum_{i=1}^{I} n_i (\bar{Y}_i - \hat{\mu})^2 / \sum_{i=1}^{I} n_i.$$
- Plugging these into the Bayes estimates yields the EB estimate (known as the *credibility formula*):
$$\hat{\mathrm{E}}[\theta_i | Y_{i1}, \cdots, Y_{i,n_i}] = \hat{\alpha}_i \bar{Y}_i + (1 - \hat{\alpha}_i)\hat{\mu},$$
where $\hat{\alpha}_i = \hat{\tau}^2 / (\hat{\tau}^2 + \hat{\sigma}_i^2 / n_i)$ is the *credibility factor* for the i^{th} class.
- An important extension, introduced by Hachemeister, is the credibility regression model that relates claim sizes to certain covariates. The credibility factor in this case has the form of a matrix.

Insurance Rate-Making: Credibility Models

- Frees, Young and Luo unified various credibility models into the framework of linear mixed models (LMM) of the form

$$Y_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + \mathbf{b}_i'\mathbf{z}_{ij} + \epsilon_{ij},$$

  with fixed effects forming the vector $\boldsymbol{\beta}$, subject-specific random effects forming the vector $\mathbf{b}_i$ s.t. $\mathrm{E}[\mathbf{b}_i] = 0$, and 0-mean random disturbances ϵ_{ij} that have variance σ^2 and are uncorrelated with the random effects and the covariates $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$.
- The credibility model $Y_{ij} = \theta_i + \epsilon_{ij}$ can be rewritten as $Y_{ij} = \beta + b_i + \epsilon_{ij}$, where $\beta = \mu$ and $b_i = \theta_i - \mu$ has mean 0 & variance τ^2.
- Estimation of $\mathbf{b}_i$ in LMM when the parameters $\boldsymbol{\beta}$ and σ_i^2 are known uses Henderson's best linear unbiased predictor (BLUP).

Evolutionary Credibility and Dynamic EB Methods

- To generalize the linear EB theory, consider longitudinal data Y_{it} for each individual i. For example, insurers data consist of claims of risk classes over successive periods.
- Frees, Young and Luo (1999) incorporated the setting of longitudinal data by replacing Y_{ij} with Y_{it} in their LMM approach; t denotes time.
- Bühlmann and Gisler (2005) further developed an evolutionary credibility theory that assumes a dynamic Bayesian model for the prior means over time.

Evolutionary Credibility and Dynamic EB Methods

- For longitudinal data $Y_{it}, 1 \le i \le n, 1 \le t \le T$, the linear Bayes estimator of the mean θ_{it} of Y_{it} assumes a prior distribution that has mean μ_t for every t. A dynamic Bayesian model specifies how μ_t evolves with time.
- One such model used in evolutionary credibility is

$$\mu_t = \rho\mu_{t-1} + (1-\rho)\mu + \eta_t,$$

  in which the η_t are i.i.d. with mean 0 and variance V.
- This is a linear state-space model, μ_t are unobserved states undergoing AR(1). μ_t can be estimated from $Y_{is}, s \le t$, by the Kalman filter $\hat{\mu}_{t|t}$ defined recursively via

$$\hat{\mu}_{t|t} = \hat{\mu}_{t|t-1} + \rho^{-1}\mathbf{K}_t(\mathbf{Y}_t - \hat{\mu}_{t|t-1}\mathbf{1}),\ \hat{\mu}_{t+1|t} = \rho\hat{\mu}_{t|t} + (1-\rho)\mu,$$

  where $\mathbf{Y}_t = (Y_{1t}, \cdots, Y_{nt})', \mathbf{1} = (1, \cdots, 1)'$ and $\mathbf{K}_t$ is the Kalman gain matrix defined recursively in terms of the hyperparameters $V = \mathrm{Var}[\eta_t], v_t = \mathrm{Var}[Y_{it}|\mu_t]$ and ρ.

Evolutionary Credibility and Dynamic EB Methods

- The Kalman filter is the minimum-variance linear estimator of μ_t. It is the Bayes estimator if $Y_{it}|\mu_t$ and η_t are normal.
- The hyperparameters μ, ρ, V and v_t in the Bayes estimate $\hat{\mu}_{t|t}$ of μ_t can be consistently estimated using the method of moments. For example, $\mu = \mathrm{E}[\mu_t]$ can be consistently estimated up to t by $\hat{\mu}(t) = (\sum_{s=1}^{t} \bar{Y}_s)/t$.
- Note that to estimate the hyperparameters, one needs the cross-sectional mean $\bar{Y}_{t-1}$ of n independent observations that have mean μ_{t-1}. An alternative approach is to replace μ_{t-1} directly by $\bar{Y}_{t-1}$, leading to

$$\mu_t = \rho \bar{Y}_{t-1} + \omega + \eta_t,$$

where $\omega = (1 - \rho)\mu$.

Linear Dynamic EB via Linear Mixed Models (LMM)

- The alternative model of μ_t leads to the LMM

$$Y_{it} = \rho \bar{Y}_{t-1} + \omega + b_i + \epsilon_{it},$$

  in which η_t is absorbed into ϵ_{it}. The random effects b_i can be estimated by BLUP.
- This is much easier to extend to nonlinear models, in contrast to the hidden Markov modeling approach that involves nonlinear filtering.
- Also, due to the form of a regression model, one can easily include additional covariates to increase the predictive power of the model in the LMM

$$Y_{it} = \rho \bar{Y}_{t-1} + a_i + \boldsymbol{\beta}' \mathbf{x}_{ij} + \mathbf{b}_i' \mathbf{z}_{ij} + \epsilon_{it},$$

  where a_i and $\mathbf{b}_i$ are subject-specific random effects, $\mathbf{x}_{it}$ represents a vector of subject-specific covariates that are available prior to time t, and $\mathbf{z}_{it}$ denotes a vector of additional covariates that are associated with $\mathbf{b}_i$.

Application to Baseball Batting Averages

- Batting average, a key performance measure in baseball, is the ratio of hits (# of successful attempts) to at bats (# of qualifying attempts).
- Efron and Morris (1975, 1977) analyzed batting averages from the first $n = 45$ at-bats of a small sample of batters in 1970 to predict their batting average for the remainder of the season.
  - Y_i and p_i denote the observed batting average and true seasonal batting average of player i, s.t. $\mathrm{E}[Y_i] = p_i$.
  - Y_i are independently distributed with $nY_i \sim \mathrm{Bin}(n, p_i)$.
  - Transformed data $X_i = n^{1/2} \arcsin(2Y_i - 1)$ for variance-stabilization.
  - Use James-Stein estimator on X_i to demonstrate the benefits of Empirical Bayes methodology.

Application to Baseball Batting Averages

- Brown (2008) analyzed batting records of Major League players over the 2005 regular season.
  - Use batting records from the 1st half season ($t = 1$) to predict the second half season ($t = 2$) performance.
  - Considered all players with at-bats $N_{it} > 10$ and have such data in both half seasons.
  - Assumed H_{it}, the number of "hits", is $\mathrm{Bin}(N_{it}, p_i)$ and used variance-stabilizing transformation

$$X_{it} = \text{arcsin} \sqrt{\frac{H_{it} + 1/4}{N_{it} + 1/2}} \sim \mathrm{N}(\text{arcsin}(p_i), \frac{1}{4N_{it}}).$$

  - Compared predictive performance of several estimators that are "motivated from empirical Bayes and hierarchical Bayes interpretations": James-Stein estimator, nonparametric EB estimator by Brown and Greenshtein (2009)

Application to Baseball Batting Averages

- Instead of a single season, use longitudinal data consisting of results from the 5 most recent seasons (2006 - 2010), or 10 half seasons $t = 1, 2, \cdots, 10$.
- Linear dynamic EB via linear mixed models (LMM)

$$X_{it} = \beta_1 \bar{X}_{t-1} + \beta_2 \bar{X}_{t-2} + b_i \ (t \geq 3),$$

  where X_{it} is same as Brown's, $\bar{X}_t$ is the average for X_{it}, b_i is the subject-specific random effects $\sim N(\alpha, \sigma^2)$.
- Training set is half seasons 3 to 9, test set is half season 10. To be comparable to Brown, require players to have both history in $t = 9, 10$ and at bats $N_{it} > 10$ for $t = 3, \cdots, 10$.
- Bayesian information criterion (BIC) selects

$$X_{it} = \beta_1 \bar{X}_{t-1} + b_i \ (t \geq 3), \ \ t = 3, \cdots, 9.$$

- Use Henderson's BLUP for one-step ahead predictions $\delta = \hat{X}_{i,10}$.

Evaluation of the Predictive Performance

- For different predictors δ of $X_{i,10}$, Brier score calculates $\sum_{i=1}^{n}(\delta - X_{i,10})^2/n$ and we also calculate the Kullback-Leibler divergence loss function (Lai, Gross, Shen 2011) given by

$$KL(\delta) = \sum_i \{Y_{i,10} \log(Y_{i,10}/\hat{p}_i(\delta)) + (1 - Y_{i,10}) \log[(1 - Y_{i,10})/(1 - \hat{p}_i(\delta))]\},$$

  where $Y_{i,10}$ is the batting average of batter i at $t = 10$, and $\hat{p}_i(\delta) = [(\sin\delta)^2(N_{i,10} + 1/2) - 1/4]/N_{i,10}$ is the predictor of $Y_{i,10}$ using δ. A smaller KL(δ) indicates better predictive performance for the group under consideration.

	LMM	Naive	Mean	EB(MM)	EB(ML)	JS
Brier	0.0045	0.0067	0.0074	0.0068	0.0060	0.0064
KL	4.45	7.03	6.90	6.32	5.68	5.99

- By making use of the longitudinal aspect of the data, the dynamic EB modeling approach implemented via LMM gives a markedly better prediction performance.

Generalized Linear Mixed Models (GLMM) & Dynamic EB

- A widely used model for longitudinal data Y_{it} in biostatistics is the generalized linear model that assumes Y_{it} with density of the form

$$f(y;\theta_{it},\phi) = \exp\{[y\theta_{it} - g(\theta_{it})]/\phi + c(y,\phi)\},$$

in which h is a smooth increasing function (the link function) and $\mathbf{x}_{it}$ is a $d-$dimensional vector of covariates s.t.

$$h(\mu_{it}) = \boldsymbol{\beta}'\mathbf{x}_{it}, \text{ where } \mu_{it} = \frac{dg}{d\theta}(\theta_{it})$$

- For the case $d = 1$ (so that $\mu_{it} = \mu_t$), Zeger and Qaqish (1988) introduced the model

$$h(\mu_t) = \sum_{j=1}^{p} \theta_j h(Y_{t-j}).$$

- Suppose the prior distribution specifies that for each $1 \leq t \leq T$, μ_{it} are i.i.d. with mean μ_t. Note that μ_s can be consistently estimated by $\bar{Y}_s$. This suggests $h(\mu_t) = \sum_{j=1}^{p} \theta_j h(\bar{Y}_{t-j})$ as an EB extension of the Zeger-Qaqish model.

Generalized Linear Mixed Models (GLMM) & Dynamic EB

- We can include fixed and random effects and other time-varying covariates of each subject i, thereby removing the dependence of $h(\mu_{it}) - h(\mu_t)$ on t in the GLMM

$$h(\mu_{it}) = \sum_{j=1}^{p} \theta_j h(\bar{Y}_{t-j}) + a_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \mathbf{b}_i'\mathbf{z}_{it},$$

  in which $\theta_1, \cdots, \theta_p$ and $\boldsymbol{\beta}$ are the fixed effects and a_i and $\mathbf{b}_i$ are subject-specific random effects.

- We assume a_i and $\mathbf{b}_i$ to be independent normal with zero means. Lai and Shih (2003) have shown by asymptotic theory and simulations that the choice of a normal distribution, with unspecified parameters, for the random effects $\mathbf{b}_i$ in GLMM is innocuous.

Generalized Linear Mixed Models (GLMM) & Dynamic EB

- Predicting the response of subject i at the next period entails estimating

$$\mu_{i,t+1} = h^{-1}(\sum_{j=1}^{p} \theta_j h(\bar{Y}_{t+1-j}) + a_i + \boldsymbol{\beta}'\mathbf{x}_{i,t+1} + \mathbf{b}_i'\mathbf{z}_{i,t+1})$$

- In general, we want to estimate some future function ψ_{t+1} of the unobserved $\mathbf{b}_i$. If we do not know $\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}$ and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_p)'$, we can estimate them by MLE using all the observations up to time t. The future value $\psi_{t+1}(\mathbf{b}_i)$ can then be estimated by

$$\hat{\psi}_{t+1,i} = \mathrm{E}_{\hat{\phi}_t, \hat{\boldsymbol{\alpha}}_t, \hat{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\theta}}_t}[\psi_{t+1}(\mathbf{b}_i) | \text{data of the } i\text{th subject up to time } t].$$

Application: Workers' compensation insurance

- The data set in Klugman (1992) contains workers' compensation losses for n = 121 occupation classes over 7 years. It relates loss to exposure (coverage), called "payroll", which is not adjusted for inflation. Also, the loss per dollar of payroll, called "pure premium", is included in the data.
- Klugman uses a variant of the credibility regression model

$$Y_{it}|(\alpha_i, \beta_i, \sigma^2) \sim N(\alpha_i + \beta_i t, \sigma^2/P_{it}),$$

in which Y_{it} is the loss of the i^{th} class in year t and P_{it} is the corresponding exposure. He reduced the effective number of parameters via the Bayesian model

$$\alpha_i|(\mu_\alpha, \tau_\alpha^2) \sim N(\mu_\alpha, \tau_\alpha^2),\ \beta_i|(\mu_\beta, \tau_\beta^2) \sim N(\mu_\beta, \tau_\beta^2),\ \mathrm{cov}(\alpha_i, \beta_i|\tau_{\alpha\beta}) = \tau_{\alpha\beta}.$$

Application: Workers' compensation insurance

- Frees, Young and Luo (2001) modified Klugman's model and applied a logarithmic transformation to the pure premium $\mathrm{PP}_{it} = Y_{it}/P_{it}$, which they used as a response variable in the LMM

$$\log \mathrm{PP}_{it} = \alpha_i + \beta_i t + P_{it}^{1/2}\epsilon_{it},$$

with $\epsilon_{it} \sim N(0, \sigma^2)$ The subject-specific variance in the above LMM is weighted by P_{it} to account for heteroskedasticity. Letting $X_{it} = \log P_{it}$, this is equivalent to

$$\log(Y_{it}) = \alpha_i + \beta_i t + X_{it} + P_{it}^{1/2}\epsilon_{it},$$

- Plotting PP_{it} (or $\log \mathrm{PP}_{it}$) versus t does not show linear trends, suggesting that inclusion of t in the model should involve random rather than fixed effects.
- Antonio and Beirlant (2006) also used year t as a covariate in evolutionary credibility. However, they used a gamma GLMM

$$Y_{it}|b_i \sim \mathrm{Gamma}(\kappa, \mu_{it}/\kappa), \;\; \log(\mu_{it}) = \alpha_i + \beta t + X_{it},$$

in which $\alpha_i \sim N(\alpha, \tau^2)$.

Application: Workers' compensation insurance

- To compare these models, we evaluate how well they predict the losses Y_{it} given the observations up to year $t-1$, for $t = 5, 6, 7$. (so the training has at least 4 years of data.)
- The 5-number summaries of the absolute prediction errors $|Y_{it} - \hat{Y}_{it}|$ for $t = 5, 6, 7$ indicates that Frees' LMM has the best overall prediction performance. This can be explained by the strong linear trend in the plot of $\log(Y_{it})$ versus $\log(P_{it})$.
- Antonio and Bierlant's GLMM performs better when the absolute errors are relatively small.
- Another important feature of the data set that has been ignored by all these models is that 7.9% of the losses are 0, and the number of zero losses tends to decrease with P_{it}.

Application: Workers' compensation insurance

- We can modify Free's LMM to allow for different slope and drop t as a regressor

$$\log Y_{it} = \alpha_i + \beta X_{it} + P_{it}^{1/2}\epsilon_{it}.$$

- To address the issue of "excess zeros", we can use a two-part GLMM:
  - Represent Y_{it} by $Y_{it} = I_{it}Z_{it}$, where $I_{it} = \mathbf{1}_{\{Y_{it}>0\}}$ and Z_{it} has the conditional distribution of Y_{it} given $Y_{it} > 0$.
  - Since $I_{it} \sim \mathrm{Bernoulli}(\pi_{it})$, we can use the GLMM

    $$\mathrm{logit}(\pi_{it}) = \rho_1 \mathrm{logit}(\bar{I}_{t-1}) + \alpha_0 + \alpha_1 X_{it} + \alpha_2 I_{i,t-1} + a_i$$

    to model π_{it}, where random effects $a_i \sim N(0, \sigma_a^2)$.
  - For $t \geq 2$, use the gamma GLMM to model the positive losses:

    $$\begin{aligned} Z_{it} &\sim \mathrm{Gamma}(\kappa, \mu_{it}/\kappa), \\ \log(\mu_{it}) &= \rho_2 \log(\bar{Z}_{t-1}) + \beta_0 + \beta_1 X_{it} + \beta_2 Z_{i,t-1} + b_i, \end{aligned}$$

    where $b_i \sim N(0, \sigma_b^2)$, $\bar{Z}_{t-1} = (\sum_{Y_{i,t-1}>0} Z_{i,t-1})/(\sum_{i=1}^n I_{i,t-1})$.

Application: Workers' compensation insurance

- We can use a hybrid model that combines the relative advantages of the modified LMM and the two-part GLMM. One way is to choose a cutoff for $X_{it} = log(P_{it})$ using its median of 17.25.
- The proposed hybrid is defined by

$$Y_{it} = \begin{cases} I_{it}Z_{it} & \text{if } X_{it} < 17.25 \\ \exp(\alpha_i + \beta X_{it} + P_{it}^{1/2}\epsilon_{it}) & \text{if } X_{it} \geq 17.25, \end{cases} \tag{1}$$

  in which $I_{it} \sim \mathrm{Bernoulli}(\pi_{it})$, $Z_{it} \sim \mathrm{Gamma}(\kappa, \mu_{it}/\kappa)$, π_{it} and μ_{it} are defined as before, $\alpha_i \sim N(\alpha, \tau^2)$, $\epsilon_{it} \sim N(0, \sigma^2)$.
- Again, select the model for each training sample (year 1 to $t-1$ for $t = 5, 6, 7$) by using BIC.

Application: Workers' compensation insurance

Table: Five-number summaries (minimum Min, 1st quartile Q_1, median Med, 3rd quartile Q_3, and maximum Max) of absolute prediction errors for different models

	LMM (Klugman)			GLMM (Antonio & Beirlant)		
	$t=5$	$t=6$	$t=7$	$t=5$	$t=6$	$t=7$
Min	717	1,168	552	282	310	207
Q_1	75,290	53,050	84,100	65,970	43,080	68,020
Med	206,800	207,800	261,200	218,000	152,900	199,600
Q_3	570,100	552,500	1,211,000	463,900	478,800	787,200
Max	20.59e6	10.70e6	10.65e6	21.19e6	8.545e6	9.943e6

	LMM (Frees)			Hybrid Model		
	$t=5$	$t=6$	$t=7$	$t=5$	$t=6$	$t=7$
Min	451	1,057	381	2	0.5	0
Q_1	67,160	35,630	41,350	52,330	43,550	43,480
Med	175,000	188,700	153,400	178,300	148,800	172,900
Q_3	572,200	535,000	455,400	630,400	513,400	415,100
Max	21.28e6	5.852e6	7.487e6	21.33e6	2.491e6	5.746e6

Baseball Batting Average Revisited

- We note that the realized batting average $Y_{it} = H_{it}/N_{it}$ is an unreliable estimate of the batter's hitting probability p_{it} when N_{it} is not large enough. Therefore Brown (2008) requires $N_{it} \geq 11$ and $N_{i,t-1} \geq 11$.
- Evaluation of the probability forecasts by Lai, Gross and Shen (2011): Estimate $m^{-1} \sum_{t=1}^{m} L(p_t, \hat{p}_t)$.
- To estimate the batter's hitting probabilities when N_{is} is small, there is even more need to rely on other batters. On the other hand, N_{is} being small may have implications on the batter's ability.
- Binomial GLMM: Random effects $b_i \sim N(\alpha, \sigma^2)$.

$$H_{it} \sim \mathrm{Bin}(N_{it}, p_{it}),\ \mathrm{logit}(p_{it}) = \beta_2 \mathrm{logit}(\bar{Y}_{t-2}) + \beta_1 \mathrm{logit}(\bar{Y}_{t-1}) + b_i.$$

- Infrequent batters: $N_{it} \leq 32 =$ 20th percentile. Brown requires $N_{it} \geq 11$ to transform to normal X_{it}.

Baseball Batting Averages for Infrequent Batters

$t = 10$	Diff Brier Loss	Diff KL Loss	Adjusted Brier
EB(MM)	800e-6	333e-5	252e-5
EB(ML)	991e-6	404e-5	271e-5
JS	848e-6	349e-5	257e-5
LMM	164e-6	227e-6	188e-5
Bin			172e-5

$t = 8$	Diff Brier Loss	Diff KL Loss	Adjusted Brier
EB(MM)	747e-6	295e-5	302e-5
EB(ML)	814e-6	322e-5	309e-5
JS	877e-6	344e-5	315e-5
LMM	394e-6	167e-5	267e-5
Bin			228e-5

$t = 6$	Diff Brier Loss	Diff KL Loss	Adjusted Brier
EB(MM)	359e-4	148e-1	348e-4
EB(ML)	429e-6	174e-5	0
JS	575e-6	239e-5	0
LMM	288e-6	138e-5	0
Bin			0

Default Modeling of Corporate Loans

- "Frailty" model for loan default: a "frailty" covariate varies over time according to an autoregressive time-series specification; using MCMC methods to perform ML estimation and to filter for the conditional distribution of the frailty process.
- Default intensity $Y_{it} = \exp(\beta_0 + \boldsymbol{\alpha}\mathbf{U}_{it} + \boldsymbol{\beta}\mathbf{V}_t + \eta F_t)$, where $\mathbf{U}_{it}$ are firm-specific covariates (Moodys distance to default, 1-year stock return) and V_t macroeconomic covariates (Treasury bill rate, 1-year return on S&P 500).
- F_t is an unobservable common economic factor "frailty") that follows an Ornstein-Uhlenbeck (continuous AR(1)) process.
- The unobservable state F_t leads to a HMM for which nonlinear filtering (via Gibbs sampler) is used to estimate F_t and MCMC is needed to estimate the parameters of the HMM (Duffie et al., 2009). EM algorithm is used to estimate the other parameters.

Default Modeling of Corporate Loans

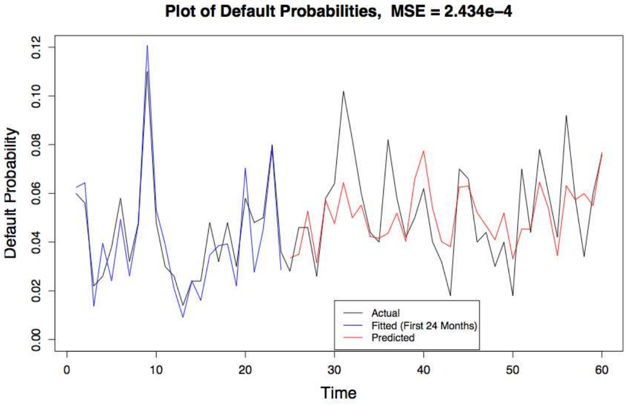
- A simpler alternative to the HMM is the proposed dynamic EB model.
- Let π_{it} denote the probability of default of firm i in the time interval $[t, t+1)$.
- We model the default indicator function Y_{it} as

$$Y_{it} \sim \mathrm{Bernoulli}(\pi_{it}),$$
$$\mathrm{logit}(\pi_{it} | Y_{i,t-1} = 0) = \rho \, \mathrm{logit}(\bar{Y}_{t-1}) + a_i + \boldsymbol{\beta}'\mathbf{U}_{it} + \mathbf{b}_i'\mathbf{V}_t,$$

  where $\bar{Y}_{t-1} = \sum_{i=1}^{n_t-1} Y_{i,t-1}/(n_t - 1)$ and a_i and $\mathbf{b}_i$ are random effects.
- This model captures the key features of Duffie's model $\lambda_{it} = \exp(\beta_0 + \boldsymbol{\alpha}\mathbf{U}_{it} + \boldsymbol{\beta}\mathbf{V}_t + \eta F_t)$ and is much simpler to implement.

Default Modeling of Corporate Loans

- Data generated from the Frailty Model of Duffie et al.; 1 month-ahead prediction. 500 companies; 24-months rolling window.



Conclusion

- We have proposed a dynamic EB model which provides flexible and computationally efficient methods for modeling panel data
- The EB approach pools the cross-sectional information over individual time series to replace an inherently complicated HMM by a much simpler GLMM.
- Replacing μ_{t-1} by the cross-sectional mean $\bar{Y}_{t-1}$ in our dynamic EB model (and thereby converting an HMM to a GLMM) is similar to using GARCH instead of SV models.
- Empirical studies in the baseball batting average and workers' compensation as well as simulation studies in corporate defaults demonstrate that our proposed model compares favorably with other models.