Technical Appendix Detailing Estimation Procedures used in 'A Flexible Approach to Modeling Ultimate Recoveries on Defaulted Loans and Bonds' (Web Supplement)

Section 1 of this Appendix provides a brief introduction to the Gibbs sampler. Section 2 clarifies the role of Gibbs sampling in the current application, Section 3 provides details of the sampling scheme used in estimation throughout the paper and Section 4 describes application of the model selection criteria.

1 Gibbs Sampling

The Gibbs sampler is an algorithm for sampling from the joint distribution of several random variables, given the distribution of each component variable conditional on all the other variables. The Gibbs sampler is commonly used in Bayesian applications to estimate the joint distribution of two or more parameters.

Consider a simple two parameter example.

Suppose we want an estimate of the joint posterior density $P(\mu, \sigma|\mathcal{D})$, where $\mathcal{D}$ is the data. The Gibbs sampler provides a means of obtaining a numerical estimate of this density if the full conditional densities $f(\mu|\sigma, \mathcal{D})$ and $f(\sigma|\mu, \mathcal{D})$ are known.

Under appropriate regularity conditions the following algorithm may be utilized:

1. Draw μ from $f(\mu|\sigma, \mathcal{D})$, conditional on some initial value of σ.
2. Draw σ from $f(\sigma|\mu, \mathcal{D})$, conditional on the drawn value of μ.
3. Go back to step 1 and draw μ conditional on the draw of σ, and continue iterating between step 1 and 2 conditioning the draw at each iteration on the previously drawn value of the other parameter.

As the algorithm proceeds and convergence is achieved, draws from the density $P(\mu, \sigma|\mathcal{D})$ is obtained.

2 The Current Application

Having specified the form of the predictive likelihood $g(y|x, \beta, h, \alpha, p)$ in equation (2), our objective is to characterize the posterior distribution of recoveries $g(y|x)$.

In principle, we could do this directly by integration:

$$g(y|x) = \int\int\int\int g(y|x, \beta, h, \alpha, p)p(d\beta, h, \alpha, p|x)d\beta dh d\alpha dp.$$

However, given the analytic intractability of the foregoing integration, we utilize the strategy of Gibbs sampling to obtain draws from $P(\beta, h, \alpha, p|x)$ as detailed in Section 3. Using the draws from $P(\beta, h, \alpha, p|x)$ to condition random draws from the distribution $g(y|x, \beta, h, \alpha, p)$ enables us to generate our sample from the target posterior distribution $g(y|x)$.

In order to implement the Gibbs sampler, we need to know the exact form of each conditional posterior distribution, namely: $P(\beta|h, \alpha, p, x)$, $P(h|\beta, \alpha, p, x)$, $P(\alpha|\beta, h, p, x)$ and $P(p|\beta, h, \alpha, x)$. By specifying proper, minimally-informative prior distributions each parameter has a well-defined conditional posterior distribution, and the parameters of the posterior are data-determined. With these considerations in mind, we employ the conditional posterior distributions detailed in Section 3.

3 The Gibbs Sampling Scheme

1. Draw from $P(\alpha|\beta, h_j, p_j, e_{ij}, y, x) \sim \mathcal{N}(\bar{\alpha}, \bar{V}_\alpha)I(\alpha_1 < \alpha_2 < \ldots < \alpha_m)$, where $\bar{V}_\alpha = \left[V_{\alpha p}^{-1} + \sum_{i=1}^{N}\left\{\sum_{j=1}^{m} e_{ij}h_j\right\} e_i e_i'\right]^{-1}$ and $\bar{\alpha} = \bar{V}_\alpha\left[V_{\alpha p}^{-1}\alpha_p + \sum_{i=1}^{N}\left\{\sum_{j=1}^{m} e_{ij}h_j\right\} e_i(y_i - x_i'\beta)\right]$,

given a Normal conjugate prior on α with mean α_p and precision $V_{\alpha p}$, subject to the labeling restriction reflected in the indicator function $I(\alpha_1 < \alpha_2 < \ldots < \alpha_m)$ that is equal to one when the prior restriction is true, and zero otherwise. The notation e_i denotes an m-vector of indicators wherein one of the elements j is equal to unity, indicating that error i is attributed to the corresponding mixture component. In specifying the prior, α_p is set equal to an m-vector of zeros and $V_{\alpha p}$ is a diagonal matrix with large numbers on the diagonal (10^5), consistent with an absence of non-sample information about the parameters of mixture components.

2. Draw from $P(\beta|\alpha, h_j, p_j, e_{ij}, y, x) \sim \mathcal{N}(\bar{\beta}, \bar{V})$, where $\bar{V} = \left[V_{\beta p}^{-1} + \sum_{i=1}^{N}\sum_{j=1}^{m} e_{ij} h_j x_i x_i'\right]^{-1}$ and $\bar{\beta} = \bar{V}\left[V_{\beta p}^{-1}\beta_p + \sum_{i=1}^{N}\sum_{j=1}^{m} e_{ij} h_j x_i (y_i - \alpha_j)\right]$, given a Normal conjugate prior β with mean β_p and precision $V_{\beta p}$. In specifying the prior, β_p is equal to a k-vector of zeros, and $V_{\beta p}$ is a diagonal matrix with large numbers on the diagonal, consistent with an absence of non-sample information about the parameters of mixture components.

3. Draw from $P(h_j|\beta, \alpha, p_j, e_{ij}, y, x) \sim \mathcal{G}(\bar{s}_j^2, \bar{v}_j)$, a Gamma distribution, where $\bar{s}_j^2 = \frac{\sum_{i=1}^{N} e_{ij}(y_i - \alpha_j - x_i\beta)^2 + v_{jp} s_{jp}^2}{\bar{v}_j}$ and $\bar{v}_j = \sum_{i=1}^{N} e_{ij} + v_{pj}$ where e_{ij} is an indicator taking a value of one if observation i is assigned to mixture component j, and zero otherwise. The prior on h_j is assumed to be Gamma, with $v_{jp} = 0.01$ and $s_{jp} = 1$.

4. Draw from $P(p_j|h_j, \beta, \alpha, e_{ij}, y, x) \sim \mathcal{D}(\bar{\rho})$, a Dirichlet distribution, where $\bar{\rho} = \left[\rho_p + \sum_{i=1}^{N} e_i\right]$ such that e_i is an m-vector of indicators wherein one of the elements j is equal to unity, indicating that error i is attributed to the corresponding mixture component. The corresponding prior vector ρ_p is an m-vector of ones.

5. Draw e_i from

$$P(e_{ij}|h_j, \beta, \alpha, p_j, y, x) \sim \mathcal{M}\left[1, \left(\frac{p_1 f_{\mathcal{N}}(y_i|\alpha_1 + x_i'\beta, h_1^{-1})}{\sum_{j=1}^{m} p_j f_{\mathcal{N}}(y_i|\alpha_j + x_i'\beta, h_j^{-1})}, \ldots, \frac{p_m f_{\mathcal{N}}(y_i|\alpha_m + x_i'\beta, h_m^{-1})}{\sum_{j=1}^{m} p_j f_{\mathcal{N}}(y_i|\alpha_j + x_i'\beta, h_j^{-1})}\right)\right],$$

a Multinomial distribution where $f_{\mathcal{N}}(.)$ denotes the normal likelihood value.

The Gibbs sampling scheme involves drawing successively from each of the distributions in steps 1-5, updating at each step the values of the conditioning parameters to be used in the subsequent draws. Looping over steps 1-5 many times results in outcomes from the joint posterior $P(\alpha, \beta, h_j, p_j, e_{ij}|y, x)$. The sampling scheme described in steps 1-5 estimates the mixture describing y conditional on x. Estimation of the joint posterior based on y alone, without conditioning on x, proceeds analogously. However step 2 is eliminated and the conditional posterior in each component of the sampling scheme is adjusted accordingly.

Unless stated otherwise, results reported in the paper are based on 10,000 iterations of the Gibbs sampler, after 100 burn-in draws are discarded.

4 Determining the Number of Mixture Components

The pooled distribution of recoveries on loans and bonds, presented in Panel (a) of Figure 1 of the paper, serves as our starting point. As noted in Section 5 of the paper, we map observations of recovery, lying between 0 and 1, to the real number line using the inverse CDF of a distribution with unbounded support on the real number line. There are many candidate distributions that we could use to perform the transformation. With no clear *a-priori* reason to prefer any of them, it would be simplest if we could state that this modeling choice is innocuous. However, it does have an impact on the shape of the transformed distribution and the subsequent choice of mixture components.

In Panel (b) of Figure 1 of the paper, we show the distribution obtained using the inverse of the Student-T CDF, with $v = 20$, to transform the data. Clearly, the transformed data remains multi modal. Using widely-applied information criteria to guide our choice of the optimal number of mixture components, it appears that the data transformed using the inverse of Student-T with $v = 20$ is optimally modeled with a 3-component mixture. As can be seen in the lower panel of Table 1, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and the Hannan-Quinn criterion (HQ) are simultaneously maximized when the number of mixture components $m = 3$. If, for example, we transform the data using the inverse of the Gaussian CDF, then the results in the upper panel of Table 1 are obtained. Not only do the information criteria suggest a larger number of mixture components, they provide inconsistent signals. That is, BIC and HQ imply the optimality of 5 mixture components, while AIC implies 6.

Table 1: Information Criteria for Varying m (Unconditional Case)

The value of each information criterion is evlauated at the posterior mean of the parameters using 10,000 draws from variants of the sampling scheme described in Appendix 3. AIC is the Akaike Information Criteria, BIC is the Bayesian Information Criteria and HQ is the Hannan-Quinn criteria. Bold type indicates preferred specification. Will add calculation details here.

	# of Mixture Components Transformation: Inverse of Standard Normal CDF				
Criteria	2	3	4	5	6
AIC	-7976	8243	8275	8319	**8325**
BIC	-8013	8187	8202	**8227**	8214
HQ	-8002	8204	8224	**8255**	8248
	# of Mixture Components Transformation: Inverse of Student-t, $v = 20$				
AIC	-9089	**7967**	7965	-	-
BIC	-9126	**7911**	7891	-	-
HQ	-9115	**7928**	7914	-	-

Given that the information criteria are best regarded as a rough guide to optimal model selection, we examine the impact of varying m on results reported in the paper. Further, in the interests of full disclosure, we note that we settled on using the transformation based on the inverse CDF of the Student-T with $v = 20$ after after estimating the model using the inverse of the Gaussian, and three variants of the Student-T: $v = 10$, $v = 20$ and $v = 50$. The information criteria obtained using $v = 20$ were consistent in suggesting a parsimonious specification with $m = 3$.