A Latent Variable Approach to Validate Credit Rating Systems using R

Chicago, April 24, 2009

Bettina Grün\textsuperscript{a}, Paul Hofmarcher\textsuperscript{a}, Kurt Hornik\textsuperscript{a}, Christoph Leitner\textsuperscript{a}, Stefan Pichler\textsuperscript{a}

\textsuperscript{a} WU Wien
Multi-rater panel provides information about rating heterogeneity

✔ across obligors,
✔ across raters,
✔ across time points.
Rating validation:
- How accurate (or precise) are PD estimates?
- Measures of rating bias and “error variance”.

Consensus PD:
- What is the consensus PD estimate when ratings differ across raters?

Explaining rating heterogeneity:
- Opaqueness of industries (Morgan 2002),
- Regional differences (Hornik et al. 2009).
Need for a probabilistic model.

Our Approach: PD treated as latent variable

Ratings are noisy observations of PD.

Provides a probabilistic model to directly estimate the rating error of a rater.

Explore patterns of heterogeneity.

Can obtain the “consensus rating” of the obligor.

Requirement: Multi rater panel with rating information on the same metric scale.

In practice: “PD ratings” (PD estimates by different raters for all obligors).
General Model I

- Merton (1974): Distance to default of firm \( i \)

\[
DD_i = a_i + b_i V_i
\]

where \( V_i \) is the log asset value of firm \( i \).

- The probability of default is given as \( PD_i = \Phi(-DD_i) \) where

\[
\Phi : (-\infty, +\infty) \rightarrow [0, 1]
\]

is the distribution function of the standard normal distribution.

- Probit of the PD is linearly related to the log asset value.
The “estimate” $PD_{ij}$ as derived by rater $j$ for the true PD of firm $i$ is

$$PD_{ij} = \Phi(-DD_i + \epsilon_{ij})$$

where $\epsilon_{ij}$ is the corresponding error (which depends on $j$ in particular).

Motivates a multi-rater model where the rater effects (“errors”) are additive on the probit ($\Phi^{-1}$, $DD_i$) scale, i.e.,

$$\Phi^{-1}(PD_{ij}) = -DD_i + \epsilon_{ij} = \Phi^{-1}(PD_i) + \epsilon_{ij}$$

Similar considerations apply for PD models based on, e.g., logistic regression.
Let \( S_{ij} = \Phi^{-1}(PD_{ij}) \) denote the score of the observed PD and \( S_i = \Phi^{-1}(PD_i) \) the latent true PD score.

Write \( \mu_{ij} \) and \( \sigma_{ij} \) for the mean and standard deviation of \( \epsilon_{ij} \), respectively.

The latent variable model can be written as

\[
S_{ij} = S_i + \mu_{ij} + \sigma_{ij} Z_{ij}
\]

where the standardized rating errors \( Z_{ij} = (\epsilon_{ij} - \mu_{ij})/\sigma_{ij} \) have mean zero and unit variance.
Specific Models I

✔ Simple model:

\[ S_{ij} = S_i + \mu_j + \sigma_j Z_{i,j} \]

✖ Rating errors are independent of obligors and their characteristics (e.g., creditworthiness).

✖ Parameters depend on raters only: \( \mu_{ij} = \mu_j, \sigma_{ij} = \sigma_j \).

✔ Extended model:

\[ S_{ij} = S_i + \mu_{g(i),j} + \sigma_{g(i),j} Z_{ij} \]

✖ Groups of obligors with homogeneous rating error characteristics.

✖ Parameters depend on rater and obligor groups:

\( \mu_{ij} = \mu_{g(i),j}, \sigma_{ij} = \sigma_{g(i),j} \).
Ensuring identifiability:

Average rating bias within each obligor group is zero.

\[ \mathbb{E}(S_{ij}) = \mathbb{E}(S_i) + \mu_{g(i),j} \text{ with } \sum_j \mu_{g,j} = 0. \]

Modelling marginal group effects:

\[ \mathbb{E}(S_i) = \nu_{g(i)} \]

\[ \text{var}(S_i) = \tau^2 \text{ or } \text{var}(S_i) = \tau_{g(i)}^2. \]
Mixed-effects model for the observed $S_{ij}$,

- latent true PD scores $S_i$ enter as random effects.

Given parametric models for the $\mu_{ij}$ and $\sigma_{ij}$ and the distributions of true PD scores and standardized rating errors.

Estimation using R:

- Marginal maximum likelihood: nlme (ratermodel), or
- Bayesian techniques: rjags and coda
Assessment of rater effects (rater heterogeneity) in terms of their $\mu/\sigma^2$ bias/variance characteristics.

Estimated random effects $\hat{S}_i$ can be interpreted as the consensus rating for obligor $i$.

Residuals: $S_{ij} - \hat{S}_i - \hat{\mu}_{ij}$, the observed scores minus the estimated random and fixed effects,

- are analyzed using some covariates (e.g., country, industry, legal form),
- are analyzed for single obligors as well as obligor groups of interest.
Data

- Data from the Austrian Credit Register from September 2007.
- Multi-rater panel containing:
  - obligor specific information (country, industry, ...),
  - rating information (original rating, PD).
- Include 13 banks which report PDs and have sufficiently many co-ratings (every bank has at least 4 co-ratings with every other bank).
  - 5432 co-ratings of 13 banks for 2090 obligors.
Variety of models analyzed using AIC and BIC.

Industry affiliation as grouping variable.

Constant variance $\tau^2$ for PD score.

$$S_{ij} = S_i + \mu_{g(i),j} + \sigma_{g(i),j} Z_{ij}, \quad S_i \sim N(\nu_{g(i)}, \tau^2),$$

$$\mathbb{E}(S_{ij}) = \mathbb{E}(S_i) + \mu_{g(i),j}, \quad \sum_j \mu_{g,j} = 0,$$

where

$\mu_{g,j}$ .... rating bias to the mean PD score of bank $j$ for obligors in industry $g$,

$\sigma_{g,j}$ .... standard deviation of the rating error of bank $j$ for obligors in industry $g$, and

$\nu_g$ ....... mean PD score in industry $g$. 
Industry specific means:

<table>
<thead>
<tr>
<th>Industry</th>
<th>$\nu_g$</th>
<th>$\Phi(\nu_g)$</th>
<th>$\Phi(\nu_g - \tau)$</th>
<th>$\Phi(\nu_g + \tau)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufac</td>
<td>-2.542</td>
<td>55.1</td>
<td>17.7</td>
<td>151.1</td>
</tr>
<tr>
<td>Energy</td>
<td>-2.993</td>
<td>13.8</td>
<td>3.8</td>
<td>44.2</td>
</tr>
<tr>
<td>Constr</td>
<td>-2.448</td>
<td>71.8</td>
<td>23.8</td>
<td>190.9</td>
</tr>
<tr>
<td>Trading</td>
<td>-2.375</td>
<td>87.7</td>
<td>29.8</td>
<td>227.5</td>
</tr>
<tr>
<td>Finance</td>
<td>-3.256</td>
<td>5.6</td>
<td>1.5</td>
<td>19.2</td>
</tr>
<tr>
<td>RealEst</td>
<td>-2.474</td>
<td>66.8</td>
<td>22.6</td>
<td>174.7</td>
</tr>
<tr>
<td>Public</td>
<td>-3.330</td>
<td>4.3</td>
<td>1.1</td>
<td>15.1</td>
</tr>
<tr>
<td>Service</td>
<td>-2.517</td>
<td>59.2</td>
<td>19.8</td>
<td>157.0</td>
</tr>
<tr>
<td>Private</td>
<td>-2.296</td>
<td>108.4</td>
<td>40.0</td>
<td>267.4</td>
</tr>
</tbody>
</table>

The corresponding standard deviation $\tau$ is 0.357 and the PD intervals are measured in basis points ($10^{-4}$).
Rating bias for rater/industry combinations:

\[ \mu_{g(i),j} \]
Rating error variance for rater/industry combinations corrected for industry effects:

\[ \sigma_{g(i),j}^2 - \sigma_{g(i)}^2 \]

[Graph showing the rating error variance for rater/industry combinations]
Residual Analysis

Across the relative exposure:

This Bank rates two obligors with high relative exposure rather too favorably.
Motivation

Formulate a statistical framework for a dynamic rating process:

- Explaining the rating heterogeneity not only cross-sectionally but also longitudinally.
- Intertemporal change of consensus ratings.
- Correlations in rating migrations (Stefanescu et al., 2009).
- Evaluation of raters.

Requirement: Multi-rater panel data with rating information on the same PD-scale.
Simple Model:

\[ S_{ij}(t) = S_i(t) + \mu_j + \sigma_j Z_{ij}(t), \]

- Rating errors are independent of obligors and their characteristics (e.g., creditworthiness).
- Parameters depend on raters only: \( \mu_{ij} = \mu_j, \sigma_{ij} = \sigma_j \).
- Latent score \( S_i(t) \) is assumed to follow an AR(1) process:

\[ S_i(t) = \alpha_i S_i(t - 1) + \nu_i (1 - \alpha_i) + \xi_i(t), \]

where \( S_i(0) = \nu_i + \xi_i(0), \nu_i \sim N(\nu, \tau^2) \) and \( \xi_i(t) \sim N(0, \phi^2) \).
- AR(1) process induces auto-correlation in the latent creditworthiness of obligor \( i \).
- Latent scores \( S_i(t) \) are uncorrelated across obligors.
Let $f(t)$ denote a market factor, latent score $S_i(t)$ can then be written as

$$S_i(t) = \alpha_i S_i(t-1) + \nu_i (1 - \alpha_i) + \beta_{1,i} f_1(t) + \beta_{2,i} f_2(t) + \xi_i(t),$$

$S_i(0) = \nu_i + \xi_i(0) + \beta_{1,i} f_1(0)$.

$\beta_{j,i} \sim N(\mu_\beta, \sigma_{\beta}^2), \ (j = 1, 2)$ denotes cross sectional correlation coefficient of scores $S_i(t)$.

Assume $f_1(t)$ follows AR(1) process

$$f_1(t) := \gamma f_1(t-1) + \zeta(t), \quad f_1(0) = \zeta(0)/\sqrt{1-\gamma^2},$$

and $f_2(t) \sim N(0, 1) \ \forall t$ to simulate market volatility shocks.
Assessment of rater effects (rater heterogeneity) in terms of their $\mu/\sigma^2$ bias/variance characteristics.

Estimated random effects $\hat{S}_i(t)$ can be interpreted as the consensus rating for obligor $i$ over time.

Market correlation structure.

Can the market explain the correlation structure between the obligors?
Bayesian Approach

- Both, prior and likelihood are used to derive the posterior
- \[ p(\theta|y) \propto p(y|\theta)p(\theta). \]

- Informative and non-informative priors \( p(\theta) \)
- Prior denotes beliefs about parameter before observing data.
- Allows for expert opinions through the use of informative priors \( \rightarrow \) informative prior.
- All information resulting in a prior arose from data \( \rightarrow \) noninformative prior.

- Gibbs sampling is used to derive posterior distribution.
The R package, rjags permits a direct interface from R to the main JAGS library (Plummer, 2008).

- `coda.samples(model, var, n.iter, thin = 5)` generates posterior sample in `mcmc.list` object.
- Function to extract random samples of the penalized deviance from jags model (`dic.samples(model, n.iter, thin = 5, type)`).
- Includes functions for model selection (`diffdic(model1, model2)`).
A JAGS model is defined by

2. The data (a list of vectors/matrices or arrays).
3. Optional, a set of initial values for the chain(s).

`jags.model()` is used to create an Bayesian graphical model

```r
> modeldyn <- jags.model(file = "model.bug", data = ratings,
                        inits = ratingsinit, n.chains = 4)
```

Compiling model graph
- Declaring variables
- Resolving undeclared variables
- Allocating nodes
- Graph Size: 23498
model
{
  f[1] <- 0;
  sf[1] <- 0;
  for (t in 2:T) {
    f[t] ~ dnorm(gamma*f[t-1], 1); sf[t] ~ dnorm(0, 1);
  }
  for (i in 1:N) {
    v[i] ~ dnorm(nu, tau);
    m[i,1] ~ dnorm(v[i] + beta[i] * f[1] + sfbeta[i] * sf[1], sigma.i);
    for (t in 2:T) {
      m[i,t] ~ dnorm(alpha[i] * m[i,t-1] + beta[i] * f[t] + sfbeta[i] * sf[t] + (1-alpha[i]) * v[i], sigma.i);
    }
    for (j in 1:J) {
      for (t in 1:T) {
        y[i,j,t] ~ dnorm(m[i,t] + mu[j], sigma[j]);
      }
    }
  }
  alpha[i] ~ dnorm(a, A); beta[i] ~ dnorm(b, B); sfbeta[i] ~ dnorm(c,C);
}
sigma.i ~ dgamma(1.0E-3, 1.0E-3);
a ~ dnorm(0, 1.0E-6)T(-1.0,1.0); A ~ dgamma(1.0E-3, 1.0E-3);
b ~ dnorm(0, 1.0E-6); B ~ dgamma(1.0E-3, 1.0E-3);
c ~ dnorm(0, 1.0E-6)T(-1.0,1.0); C ~ dgamma(1.0E-3, 1.0E-3);
nu ~ dnorm(0, 1.0E-6); tau ~ dgamma(1.0E-3, 1.0E-3);
gamma ~ dnorm(0, 1.0E-6)T(-1.0,1.0);
Output diagnostic

- Output diagnostic can be done with R package coda (Convergence Diagnosis and Output Analysis, Plummer et al., 2008).
- Coda package includes functions for convergence tests of MCMC chains.
  - autocorr(x, lags = c(), ) calculates the autocorrelation for the MCMC chain (x = mcmc.obj) at the lags lags.
  - gelman.diag(x, ) calculates the scale reduction factor for a variable x.
  - geweke.diag(x, ) can be used for Geweke diagnostic.
Data

✔ External monthly rating data from the Big Three rating agencies Standard & Poor’s, Fitch and Moody’s.

✔ Time period: February 2007 to January 2009

✔ This study: 87 corporates out of the 125 corporates included in the iTraxx Europe index (Series 10)

✘ iTraxx: most-liquid CDS referencing European investment-grade entities.

✔ Source: Historical rating announcements taken from Reuters Credit Views.
Mapping of ordinal ratings

- ✔ Rating agencies provide ratings only on a ordinal scale
  - ✘ Unclear whether PD or EL is measured.
  - ✘ Unclear which time horizon.
  - ✔ Long history of default rates available.
  - ✔ However, agency ratings are widely used as PD-equivalents by the industry (Erlenmaier, 2006; Tasche, 2008).
  - ✔ This study: Mapping of ordinal ratings to PD provided by OENB.

- ✔ Note: All results of this study are conditional on used mapping procedure.
  - ✘ Quality of rating system can not be separated from mapping (in particular bias).
## Model Selection

<table>
<thead>
<tr>
<th>Model Selection</th>
<th>Variety of models analyzed using deviance information criteria.</th>
<th>✔</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static Model</td>
<td>Simple dynamic model.</td>
<td>✗</td>
</tr>
<tr>
<td>Empirical Example I</td>
<td>Dynamic market model including the EUROSTOXX50 as an exogenous market factor.</td>
<td>✗</td>
</tr>
<tr>
<td>Dynamic Model</td>
<td>Dynamic market model with a market factor following an AR(1) process.</td>
<td>✗</td>
</tr>
<tr>
<td>Estimation</td>
<td>Dynamic market model with $f(t) = (f_1(t), f_2(t))$ with $f_1(t)$ as an AR(1) process and $f_2(t) \sim N(0, 1)$</td>
<td>✔</td>
</tr>
<tr>
<td>Empirical Example II</td>
<td>Dynamic market model with two market factors performed best.</td>
<td>✔</td>
</tr>
<tr>
<td>Data Mapping</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Convergence Diagnostic – Trace Plot

✓ **Trace plot** for parameter $\mu_\alpha$.

✘ Trace plot is plot of the iteration number against the value of drawn of parameter for each iteration

✘ **Burn-in** period, about 5000 iterations

![Trace plot](image-url)
Convergence Diagnostic – Gelman Rubin Diagnosis

✔ Calculate the estimated variance of a parameter as a weighted sum of the within-chain $W$ and between-chain variance $B$.

✗ Gelman Rubin factor $R$ is a ratio of $W$ and $B$. 

![Graph showing the Gelman Rubin factor over iterations](image)
Rating Errors

Rating bias and rating error variance for the three raters:

<table>
<thead>
<tr>
<th></th>
<th>Fitch</th>
<th>Moody’s</th>
<th>S&amp;P</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.0298 (0.55 bp)</td>
<td>-0.1085 (-1.59 bp)</td>
<td>0.0787 (1.57 bp)</td>
</tr>
<tr>
<td>SD</td>
<td>0.0031</td>
<td>0.0044</td>
<td>0.0024</td>
</tr>
</tbody>
</table>

Bps relative to the grand mean $\nu (-3.295, \text{ in bps: 5.01})$.

<table>
<thead>
<tr>
<th></th>
<th>Fitch</th>
<th>Moody’s</th>
<th>S&amp;P</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.1443</td>
<td>0.1987</td>
<td>0.0156</td>
</tr>
<tr>
<td>SD</td>
<td>0.0025</td>
<td>0.0045</td>
<td>6e-04</td>
</tr>
</tbody>
</table>
Consensus II

Enel SpA

Time

PD (in bp)

0 2 4 6 8

07−02 07−05 07−08 07−11 08−02 08−05 08−08 08−11

Fitch  
Moody's  
S&P  
Consensus

Consensus II
Market Factor

Motivation

Static Model

Empirical Example I

Dynamic Model

Estimation

Empirical Example II

Data

Mapping

Model Selection

Convergence

Diagnostic I

Convergence

Diagnostic II

Rating Errors

Consensus I

Consensus II

Market Factor

Summary

Indices (in 1000 points)

f
EUR50
DAX
FTSE100
CAC40

Time

07-02 07-05 07-08 07-11 08-02 08-05 08-08 08-11
Summary

✓ This methodology can be used to make fundamental statements about the heterogeneity of raters of a multi-rater panel.

✓ The “consensus rating” of obligors can be estimated.

✓ Using the consensus rating a residual analysis can be performed for single obligors as well as obligor groups of interest.

✓ The effect of a market factor on the consensus can be estimated.

Requirement: The number of co-ratings within a subgroup must be sufficiently large.