Estimating Default Probability

In this lecture. . .

- Cross-asset impact of default probability   Mkt sources
- Generalized Linear Models (GLM): a likelihood approach to estimation and inference   Altman Z-score
- Estimation of default probability for an enterprise with logit and probit regressions
- Sovereign credit rating transitions with the ordered probit

By the end of this lecture you will be able to

- Understand sources of default probability information.
- Apply a generalized linear model in a multivariate setting– that is, perform estimation by logit and probit regressions.
- Conduct inference with Maximum Likelihood Estimation: analyse robustness of estimates and test for significance.
- Understand credit ratings migration.

Purpose

Probability of default (PD)

- Bootstrapped from the market data (CDS, risky bonds) and used for CVA and risk calculations as well as capital structure arbitrage.

- Can be estimated from historical data by statistical methods (logit or probit regression) and used for credit migration and other ratings analytics.

regression

Altman Z score

GLM (probit, logit)

Probability is dependent variable.

Capital structure arbitrage

While, **capital structure arbitrage** term is traditionally used for trading of equity against convertible bonds, it also covers the arbitrage with market-traded credit spreads (CDS).

One can compare PD estimations from different market sources in order to identify rich and cheap claims.

- For example, PD bootstrapped from term structure of credit spreads can be compared to ones implied by a risky bond curve (spot curve).

$1 - \Pr \text{Surv} = PD \Rightarrow$ Hazard Rates via Poisson process dist

$p(0,T)$

dm

Credit Triangle

Credit Triangle rule of thumb suggests that CDS is proportional to default intensity (hazard rate). This is known as

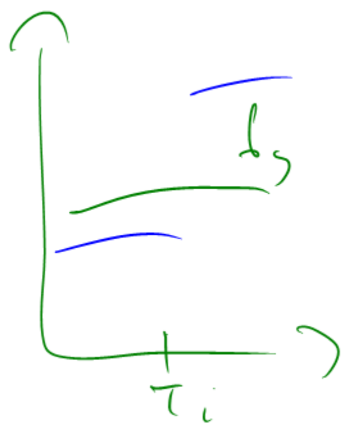
$$\underset{10\text{ bps}}{\text{CDS}} \approx \lambda(1 - \underset{0.4}{RR}).$$

$\lambda dt = p\, dt$

To obtain the term structure of piecewise constant hazard rates you can use the bootstrapping of survival probabilities from CDS (JPM formulation) and the following relationship:

$$\underset{\text{Pr Surv}}{P(0,T)} = e^{-\int \lambda_s ds} = \underbrace{e^{-\int_0^{T_i} \lambda_s ds}}\times e^{-\int_{T_i}^{T} \lambda_s ds} \quad (1)$$

and



$$\lambda_i = -\frac{1}{\tau} \log \frac{P(0,T_i)}{\underbrace{P(0,T_{i-1})}} \quad (2)$$

We can express the intensity as a ratio of survival probabilities.

Risky Bond

Remember how bond pricing equation (BPE) derivations of Intensity Models lecture added a spread p to the short rate $r(t)$.

$$Z(r, p, t; T) = \mathbb{E}^{\mathbb{Q}}\left[e^{-\int_t^T (r_s + p_s)ds} | \mathcal{F}_t\right] \qquad r(t) \tag{3}$$

$pdt = \lambda dt \in$ CDS

Over a small time $pdt \approx \lambda dt$, piecewise constant assumption about intensity gives

$$\int_0^T p_s ds = \int_0^T \lambda_s ds = \lambda_1\tau_1 + \lambda_2\tau_2 + \ldots + \lambda_n\tau_n = \lambda_T\tau$$

But $\lambda(t)$ can itself be a function or stochastic process!

IMP

Volatility Skew for Equity Options

What are the consequences of using the risky rate instead of risk-free rate in the Black-Scholes equation?

$$r \to (r + p)$$

By doing so, we obtain the Merton Model!

Adding a small credit spread p to a risk-free rate r in the pricing PDE induces a skew though of less magnitude than observed in the markets.

- Empirical evidence of correlation between credit spreads (5Y), implied volatility, and volatility skew

$$CDS = a + b\,\sigma_{ATM} + c\ SKEW$$

$$\Delta = N(d_1)$$

$$SKEW = \sigma_{\Delta 0.25} - \sigma_{\Delta 0.5}$$

ATM

Market Sources of Default Probability

We had a quick overview of how default probability transpires in models for bonds and equity options.

A brief methodology list for market sources:

- credit default swaps (by bootstrapping IHP) CDS Lecture
- risky bond prices (by approach from Hull and White) Z_I / Z_0
- equity option prices (by skew, extending Merton Model) $(r+p)$

Let's turn to the forth method of estimating the probability of default statistically.

Statistical Estimation for Probability of Default

Linear regression model

The first call to model a relationship between the response variable and its explanatory variables is a linear regression.

$\mathbb{E}[Y|X]$

$$\boldsymbol{Y} = \boldsymbol{\beta X} + \boldsymbol{\epsilon}$$

Dependent Explanatory (Covariate)

- The assumption of residuals being *iid* Normal goes into construction of Maximum Likelihood

$$\epsilon_t \sim N(0, \sigma^2)$$

- Coefficients $\hat{\boldsymbol{\beta}}$ are such that maximise the joint likelihood for all observations, where each residual ϵ_t is conditionally independent.

Estimating PD with a regression

- If we would like $\mathbf{Y}$ to directly give PD for each name then

$$\sum \boldsymbol{\beta} \boldsymbol{X_i} \in [0, 1]$$

- Response variable $\boldsymbol{Y}$ can be an indicator (default/no default) or ordinal (rating), implying Bernoulli or Binomial probability density respectively.

y {0,1} ⇒ implied PD by βX_i

- The relationship between $\boldsymbol{Y}$ and PD might be non-linear therefore, requiring a link function $\boldsymbol{Y} = g(\boldsymbol{p})$.

The simple linear regression is **not** suited to model PD.

Default event

Default is a *response variable* modelled from a few explanatory or *independent variables* that represent credibility of a debtor.

Default is a **binary** variable.

$$Y = \begin{cases} 1 & \text{default} \\ 0 & \text{no default} \end{cases}$$

Probability of default $p = \mathbb{E}[Y]$ is calculated as a frequency.

$$PD = \frac{\sum N_{Y=1}}{N}$$

That is an average number describing the population but **not a model** that gives a prediction.

Conditional expectation

$$p = \mathbb{E}[\boldsymbol{Y}|\boldsymbol{X}]$$

Why conditional?

- It allows us to model default events y_i as **independent**.

- Combinations of independent events (defaults) are modelled by the Binomial Distribution.

Multivariate GLM: Covariate X

1. Using a set of k explanatory variables, called the *covariate* $\mathbf{X}$

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \cdots & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{pmatrix}$$

GLM is estimated under expectation $PD = \mathbb{E}[\mathbf{Y}|\mathbf{X}] = \boldsymbol{X}\boldsymbol{\beta'}$

$$\begin{array}{c} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} \quad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \cdots & \vdots \\ x_{N1} & \cdots & x_{Nk} \end{pmatrix}}_{\text{financial ratios}} \times \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

Altman Z-score Model

By discriminant analysis of variance, Altman Z-score identifies the following factors from a set of 22 variables

X_1 - Working Capital/Total Assets

X_2 - Retained Earnings/Total Assets

X_3 - Earnings Before Interest and Tax/Total Assets

X_4 - Market Value of Equity/Total Liability (by book value)

X_5 - Sales/Total Assets

Non-manufacturing and non-US samples can lead to noticeably different $\boldsymbol{X}\boldsymbol{\beta}'$ than the original estimation.

Altman Z-score in GLM Framework

The original model was estimated as

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.999X_5$$

Normal, Z-score

The model of this kind is a **probit regression** that requires independent variables X_i to be Normally distributed (Z-scores).

$$\hat{\boldsymbol{\beta}}\boldsymbol{X} \sim \text{Normal} \qquad \text{gives} \qquad \text{PD or } p = \Phi(\boldsymbol{X}\boldsymbol{\beta}') = \Phi(Z)$$

[0, 1]

What if $|\boldsymbol{\beta}| > 1$? Then we can't use the probit model because $\boldsymbol{X}\boldsymbol{\beta}'$ will not conform to probability density.

Non-linear link

We noted that for the probit model, the link is inverse of Normal *cdf*. How so?

$\phi^{-1}(\phi(Y)) = Y$

Probit

$$
\begin{aligned}
p &= \Phi(\boldsymbol{X}\boldsymbol{\beta}') = \Phi(\boldsymbol{Y}) \\
\Phi^{-1}(p) &= \Phi^{-1}(\Phi(\boldsymbol{Y})) \quad \text{so} \\
Y = g(p) &= \Phi^{-1}(p)
\end{aligned}
$$

For the linear regression model,

GLM

$$
\begin{aligned}
\mathbf{Y} &= \boldsymbol{X}\boldsymbol{\beta}' + \boldsymbol{\epsilon} \\
PD \equiv \mathbb{E}[\mathbf{Y}|\mathbf{X}] &= \boldsymbol{X}\boldsymbol{\beta}' \\
g(p) &= \boldsymbol{X}\boldsymbol{\beta}' \qquad \text{and} \\
p &= g(\boldsymbol{X}\boldsymbol{\beta}')^{-1}
\end{aligned}
$$

inverse link

Probability of default p comes as a latent variable.

Link function generalises the regression

For a Binomial density, including binary default event $y_i = \{1, 0\}$ and ordinal ratings $y_i = 1, 2, 3, 4, 5$, the inverse of any *cdf* can be used as a link function.

Linear part $\boldsymbol{\beta X}$ is linked to probability of default by $p = g(\boldsymbol{X\beta'})^{-1}$.

The link itself is non-linear but the function $g(p)$ must be differentiable and monotonic.

Response variable $\mathbf{Y}$ might have any density, and inputs $\boldsymbol{X}$ do not have to be Normal variables.

Multivariate GLM: Link function

2. A **link function** is the clever bit allowing convert a default event indicator $y_i = \{1, 0\}$ to the probability p_i

$$p_i = g^{-1}(\boldsymbol{X_i}\boldsymbol{\beta'})$$ inverse of link $g(p)$

$$\begin{array}{c} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{array} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \Rightarrow \begin{pmatrix} g(\boldsymbol{X_1}\boldsymbol{\beta'})^{-1} \\ \vdots \\ g(\boldsymbol{X_n}\boldsymbol{\beta'})^{-1} \end{pmatrix} \Rightarrow \begin{pmatrix} PD_1 \\ \vdots \\ PD_n \end{pmatrix} \begin{array}{c} 0.07 \\ 0.03 \end{array}$$

Notice that we obtain PD by a model $\boldsymbol{X}\boldsymbol{\beta'}$, for which $\hat{\boldsymbol{\beta}}$ have <u>to be estimated</u>.

Neural net

Regressions for categorical (binary) response variable $y_i = \{1, 0\}$ (logit and probit) are a case of **a neural net** modelling, known as a single-layer *perceptron*.

We process multiple inputs $\boldsymbol{X_i}$ into one implied PD, which is transformed into prediction $\widehat{y}_i$ (default/no default)

$$
\begin{array}{ccccc}
X_{i,1} \rightarrow & & & & \\
\vdots & & & & \\
X_{i,2} \rightarrow & PD_i & \rightarrow & \widehat{y}_i \\
\vdots & & & & \\
X_{i,k} \rightarrow & & & &
\end{array}
$$

If PD_i is above threshold a neuron 'fires' output: $\widehat{y}_i = 1$. Neuron is modelled with the logistic step function.

Towards Maximum Likelihood

A regression model is estimated by maximising over the log-likelihood function $\log \boldsymbol{L}$

For the linear regression, maximum likelihood analytical solutions for $\boldsymbol{\beta}$ are known.

Let's start working towards the expression for Maximum Likelihood for GLM, specifically for the default event variable

$$Y = \begin{cases} 1 & \text{default} \\ 0 & \text{no default} \end{cases}$$

Bernoulli Variables

Each default/no default outcome (Bernoulli draw) has a set of its own explanatory variables $\boldsymbol{X_i}$.

$$y_i|\boldsymbol{X_i} \sim \text{Bernoulli}(p_i)$$

$$\mathbb{E}\left[y_i|\boldsymbol{X_i}\right] = p_i$$

$$\Pr\left(y_i = 1, 0\right) = \begin{cases} p_i \\ 1 - p_i \end{cases}$$

$$\Pr\left(y_i = 1, 0\right) = p_i^{y_i}(1 - p_i)^{1-y_i}$$

Each outcome is determined by an **unobserved** probability of default.

Log-likelihood

for Bernoulli: (not Normal)

We begin with Bernoulli density for a single observation

{0, 1} {1, 0}

$$f(y_i; p_i) = p_i^{y_i}(1-p_i)^{1-y_i}$$

Single

Its contribution to the log-likelihood is

$$\log f(y_i; p_i) = y_i \log p_i + (1-y_i)\log(1-p_i)$$

The joint log-likelihood for multiple default events $y = \{1, 0\}$ observed together is given by

$$\begin{aligned} \log f(y_1, y_2, \ldots, y_N) &= \log \prod_{i=1}^{N_{obs}} f(y_i; p_i) \\ &= \sum_{i=1}^{N_{obs}} \log f(y_i; p_i) \end{aligned}$$

multiple

Joint log-likelihood

$$\log L = \sum_{i=1}^{N_{obs}} \left[\underbrace{y_i}_{\text{Data}} \log \overbrace{p_i}^{\Lambda(X\beta')} + (1 - \underbrace{y_i}_{\text{Data}}) \log(1 - p_i) \right] \tag{4}$$

y_i is known from dataset. Default probability p_i comes from the regression model $\boldsymbol{X_i \beta'}$ but we need a link function

$$p_i = g^{-1}(\boldsymbol{X_i \beta'})$$

To understand *which specific link function* to use we have to consider canonical form of Bernoulli density as a member of the Exponential Family of distributions.

Bernoulli Density

We can express Bernoulli density for a random variable $y = \{1, 0\}$

$$f(y;p) = p^y(1-p)^{1-y} = \exp\left[y \log\left(\frac{p}{1-p}\right) + \log(1-p)\right]$$

Choice of a link function is the same for any categorical $\boldsymbol{Y}$

$$\boxed{g(p) = \log\left(\frac{p}{1-p}\right)}$$

This is a **logit function**, which can be read as the **log of odds**.

$g^{-1}(\ )$ logistic function

Logit Model

Relating the logit function to the linear regression gives

$$
\begin{aligned}
g(p) &= \boldsymbol{X}\boldsymbol{\beta'} \\
\log\left(\frac{p}{1-p}\right) &= \boldsymbol{X}\boldsymbol{\beta'} \\
p/(1-p) &= e^{\boldsymbol{X}\boldsymbol{\beta'}}
\end{aligned}
$$

Also remember that

$$p = g(\boldsymbol{X}\boldsymbol{\beta'})^{-1}$$

it is possible to deduce that

$$p = g(g(p))^{-1}$$

Logit Model

Result for the probability of default gives logistic function

$$p = \frac{e^g}{1+e^g} = \frac{1}{1+e^{-g}}$$

In linear model terms for $\boldsymbol{X\beta'}$ we have a **logistic regression**

$$\boxed{p = \frac{e^{\boldsymbol{X\beta'}}}{1+e^{\boldsymbol{X\beta'}}}}$$

$g(p) = X\beta'$

$$\Lambda(\boldsymbol{X\beta'}) = \frac{\exp(\boldsymbol{X\beta'})}{1+\exp(\boldsymbol{X\beta'})}$$

We defined the term to insert for p_i in the log-likelihood function $\log L$, so that it reflects a regression model.

Log-likelihood of Logit Model

The likelihood of a logistic regression uses joint likelihood (4).

$$\log \boldsymbol{L} = \sum_{i=1}^{N_{obs}} \Big[\underset{\uparrow \text{Data}}{y_i} \log \Big(\Lambda(\underset{\uparrow}{X_i} \boldsymbol{\beta}') \Big) + (1 - \underset{\uparrow \text{Data}}{y_i}) \log \Big(1 - \Lambda(\underset{\uparrow}{X_i} \boldsymbol{\beta}') \Big) \Big] \quad (5)$$

We are ready to set up this expression in Excel and run a numerical Solver that varies $\hat{\boldsymbol{\beta}}$ until the function is maximised.

Each observation (row), $\boldsymbol{X}_i$ gives a prediction for $p_i = \Lambda(\boldsymbol{X_i}\boldsymbol{\beta}')$ which we can compare to the realised outcome $y_i = 1, 0$, e.g., default/no default.

Log-likelihood of Logit Model

Analytical solution to the optimisation task

$$\underset{\boldsymbol{\beta}}{\operatorname{argmax}} \log \boldsymbol{L}$$

is tasking and would require finding solutions for $\hat{\boldsymbol{\beta}}$ by setting derivatives to zero

$$\frac{\partial \log \boldsymbol{L}}{\partial \beta_j} = 0 \quad , \ldots , \quad \frac{\partial \log \boldsymbol{L}}{\partial \beta_k} = 0$$

3. To build an expression for Maximum Likelihood (over which we optimise) we need an explicit distribution for y_i.

$$y_i|\boldsymbol{X_i} \quad \sim \quad \text{Bernoulli}(p_i)$$

$$p_i \quad \text{is} \quad \Pr(y_i = 1, 0|\boldsymbol{X_i})$$

$$\Pr\left(y_i = 1, 0\right) \quad = \quad \begin{cases} p_i \\ 1 - p_i \end{cases}$$

Notice that response variables y_i are independent but **not identically distributed**. Each outcome has its specific p_i.

"Each observed company has its own probability of default in a given year."

Exponential Family

Most of the familiar distributions belong to the EF: Normal, Chi-squared, Binomial, Poisson, Gamma, Beta... **not** Student's t.

$$f(y;\theta) = e^{a(y)b(\theta)+c(\theta)+d(y)}$$

Expressing a *pdf* in *canonical form* requires one parameter only, $\theta = \mathbb{E}[y]$.

For the Normal distribution $\theta = \mu$, for Binomial $\theta = p$.

The variance for any of the Exponential family's distribution can be expressed as

$$\mathbb{V}ar[y] = V(\mathbb{E}[y])\,\phi.$$

MLE Summary

Each y_i can follow any of the Exponential Family distributions. That achieves realistic representation of variability in $y_1, \ldots, y_N$.

$$\mathbb{E}[y_i|\boldsymbol{X}_i]) = g(\boldsymbol{X_i\beta'})^{-1} = p_i$$

$$y_i \sim EF(g(\boldsymbol{X_i\beta'})^{-1}, \phi) \quad \phi = 1$$

The twist is that we conduct MLE to estimate k parameters $\widehat{\boldsymbol{\beta}}' = [\beta_1, \beta_2, \ldots, \beta_k]$ **not** N values of p_i directly.

Because of the known result for $\mathbb{V}ar[y]$, we can be mistaken about the distribution of y_i but still construct a likelihood function and obtain acceptable $\hat{\boldsymbol{\beta}}$. This is called **Quasi-MLE**.

Altman Z-score model replication using **logit**

Implementation in Excel...

These estimates were obtained by likelihood maximisation for logistic regression.

$\hat{\beta}$

J

	A	B	C	D	E	G	H
1	**Model 1. Altman Z-score using logit**						
2	Y	Default indicator				Parameter	Estimate
3	X0	Const				C	-2.54
4	X1	Working capital/Total Assets				Beta1	0.41
5	X2	Retained Earnings/Total Assets				Beta2	-1.45
6	X3	Earnings Before Interest and Tax/Total Assets				Beta3	-8.00
7	X4	Market Value of Equity/Total Liability				Beta4	-1.59
8	X5	Sales/Total Assets				Beta5	0.62
9							
10							
11	**Model 2. Restricted (by significant coefficients)**						
12	Y	Default indicator				Parameter	Estimate
13	X0	Const				C	-2.32
14	X2	Retained Earnings/Total Assets				Beta2	-1.42
15	X3	Earnings Before Interest and Tax/Total Assets				Beta3	-7.18
16	X4	Market Value of Equity/Total Liability				Beta4	-1.62

Log L_{M1}

Log L_{M2}

Implementation in Excel: Data

The dataset consists of 5 ratios $\boldsymbol{X_i}$ and the binary default event $y_i = \{1, 0\}$ recorded for 830 firms over 6 years.

	A	B	C	D	E	F	G	H	I
1	Firm ID	Year	Default, Y	Const	WC/TA	RE/TA	EBIT/TA	ME/TL	S/TA
2	1	1999	0	1	0.501	0.307	0.043	0.956	0.335
3	1	2000	0	1	0.55	0.32	0.05	1.06	0.33
4	1	2001	0	1	0.45	0.23	0.03	0.80	0.25
5	1	2002	0	1	0.31	0.19	0.03	0.39	0.25
6	1	2003	0	1	0.45	0.22	0.03	0.79	0.28
7	1	2004	0	1	0.46	0.22	0.03	1.29	0.32
8	2	1999	0	1	0.01	-0.03	0.01	0.11	0.25
9	2	2000	0	1	-0.11	-0.12	0.03	0.15	0.32
10	2	2001	0	1	0.06	-0.11	0.04	0.41	0.29
11	2	2002	0	1	0.05	-0.09	0.05	0.25	0.34
12	2	2003	0	1	0.12	-0.11	0.04	0.46	0.31
13	3	1999	0	1	-0.04	0.27	0.05	0.59	0.21
14	3	2000	0	1	-0.04	0.25	0.03	0.33	0.21
15	3	2001	0	1	0.00	0.15	0.00	0.16	0.16
16	3	2002	0	1	-0.05	0.02	0.01	0.07	0.16
17	3	2003	0	1	-0.03	-0.01	0.02	0.10	0.18
18	3	2004	0	1	-0.03	-0.04	0.02	0.09	0.19

Implementation in Excel: Logistic Link

βX_i p

		beta*X	PD	Log L
Nobs	4000	-4.45	1.16%	-0.0117
Population PD	1.80%	-4.69	0.91%	-0.0092
		-4.03	1.75%	-0.0177
		[illegible]	[illegible]	-0.0330
		-4.02	1.76%	-0.0177
		-4.79	0.82%	-0.0083
		-2.57	7.10%	-0.0736
Null Hypothesis		-2.71	6.25%	-0.0645
C	-4.00	-3.15	4.09%	-0.0418
Beta1	0	-2.95	4.97%	-0.0510
Beta2	0	-3.16	4.07%	-0.0415
Beta3	0	-4.13	1.58%	-0.0160
Beta4	0	-3.57	2.74%	-0.0277
Beta5	0	-2.90	5.22%	-0.0536

=C2*LN(logistic(M2))+(1-C2)*LN(1-logistic(M2))

=logit(L3)

Notes:

1. Y={1,0} default idicator
Population PD is an average of Y

2. For each observation

$$p_i = \Lambda(\boldsymbol{X_i}\boldsymbol{\beta}')$$

3. Likelihood Maximisation Magic

Solver varies estimates $\hat{\beta}$
to maximise the sum of log-likelihoods
(probability mass)

The population PD converted $\Lambda(1.80) = 4.00$ provides an intercept for this logistic regression.

Implementation in Excel: MLE Setup

For each observation (row) $\boldsymbol{X_i \beta'}$ is calculated and then, converted to default probability by the logistic function

$$p_i = \Lambda(\boldsymbol{X_i \beta'})$$

Contribution to the likelihood from each observation is

$$\begin{aligned} \log L_i &= y_i \log \left[\Lambda(X_i \beta') \right] + (1 - y_i) \log \left[1 - \Lambda(X_i \beta') \right] \\ &= y_i \log p_i + (1 - y_i) \log (1 - p_i) \end{aligned}$$

$\Sigma \, \text{Log} \, L_i = \text{Log} \, L^{*}$

The probability mass (sum of all likelihoods)

$$\ell_{\mathbf{Y}} = \sum_{i=1}^{N_{obs}} L_i$$

Contributions are added up and the total is **maximised**.

We started with any assumed $\hat{\beta}$ and run Solver to find the regression coefficients that maximise the total likelihood.

	Estimates
C	-2.543
Beta1	0.414
Beta2	-1.454
Beta3	-7.999
Beta4	-1.594
Beta5	0.620

Sum log L
-280.526

M1

	Estimates
C	-2.318
Beta2	-1.420
Beta3	-7.179
Beta4	-1.616

Sum log L
-282.219

M2 restricted

Two models presented here, the second is a restricted model, re-estimated after with insignificant coefficients (variables) were excluded.

MLE properties

If we use a correct distribution of the dependent variable $\boldsymbol{Y}$ to construct the Likelihood function, estimation has nice properties

1. **Efficient** – the estimates $\hat{\boldsymbol{\beta}}$ have the smallest variance

2. **Consistent** – as sample size gets large, this becomes small

$$\Pr(|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}| > \text{ Tolerance})$$

   When an attempt is made to characterise MLE, you will see this difference and proofs about which distribution it follows.

MLE Analysis: asymptotic efficiency

For the observations y_i drawn from the Exponential Family, the regression estimates asymptotically converge

$$\hat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \mathbf{I}^{-1}\right)$$

We don't know the true $\boldsymbol{\beta}$ but it is possible to calculate the Information Matrix. It's inverse $\mathbf{I}^{-1}$ provides the standard errors.

We hear about the information matrix for the first time. But let's use the stylised example of Normal Distribution to show

$$\mathbf{I} = -\mathbb{E}\left[\frac{\partial^2 \log \mathbf{L}}{\partial\mu\partial\mu}\right] = \frac{T}{\sigma^2}.$$

S.O. Partial
wrt distribution
parameter

(see Workings).

$$\mathbf{I}^{-1} = \sigma^2/T$$

The inverse of information matrix is an easily recognised as **the standard error** (squared). If $T \to \infty$ there is no estimation error!

- For the large samples, we say MLE is asymptotically efficient: the standard error of estimates $\hat{\boldsymbol{\beta}}$ is minimised (we also say "estimates are robust").

- There is no way to tell if a particular small sample provides biased estimates (or not) *wrt* the unknown true estimates $\boldsymbol{\beta}$.

Implementation in Excel: Information Matrix

The diagonal of the inverse of **Information Matrix**, the $\mathbf{I}^{-1}$, provides the squared standard errors for regression coefficients.

Inverse of Information

C	Beta1	Beta2	Beta3	Beta4	Beta5
0.07	-0.02	0.02	-0.16	-0.06	-0.04
-0.02	0.33	-0.03	-0.14	-0.02	-0.02
0.02	-0.03	0.05	0.01	-0.02	0.00
-0.16	-0.14	0.01	7.30	-0.05	-0.13
-0.06	-0.02	-0.02	-0.05	0.10	0.01
-0.04	-0.02	0.00	-0.13	0.01	0.12

Inference Table

Paramete	Estimates	Std err	t-stats
C	-2.54	0.27	-9.56
Beta1	0.41	0.57	0.72
Beta2	-1.45	0.23	-6.34
Beta3	-8.00	2.70	-2.96
Beta4	-1.59	0.32	-4.93
Beta5	0.62	0.35	1.77

Information Matrix in GLM (analytical solution)

For probability of default, we estimate **the logistic regression** over the binary (categorical) response variable $y_i = 1, 0$.

$$\mathbf{I} = -\,\mathbb{E}\left[\frac{\partial^2 \log \mathbf{L}}{\partial\beta_j \partial\beta_k}\right] = \mathbf{X}\mathbf{P}'\mathbf{X}$$

The information matrix is a **Hessian** (second-order derivative) over the log-likelihood. $\mathbf{P}$ is a diagonal matrix of $p_i(1 - p_i)$. Computationally, each element of information matrix is

$$\mathbf{I}_{j,k} = \sum_{i=1}^{N_{obs}} p_{i,j}(1 - p_{i,j}) x_{i,j} x_{i,k} \tag{6}$$

This is how the Information Matrix is computed in VBA code.

A few observations can be made about the calculation of Information Matrix $\mathbf{I}$ in Equation (6)

- It is done using MLE-estimated $\hat{\boldsymbol{\beta}}$ because we get fitted PD

$$p_i = \Lambda(X_i'\hat{\boldsymbol{\beta}})$$

- $p_i(1-p_i)$ is a contribution to Binomial variance.

- Information Matrix $\mathbf{I}$ does **not** depend on y_i (its distribution).

  The standard errors (significance) does not depend on how we specify the density of response variable $\mathbf{Y}$.

Model Selection

Within the regression, certain variables come up as insignificant.

We would like to check if a more compact model is just as likely to deliver the same likelihood.

In our Altman Z-score study, the following variables come across as insignificant to the model (look at p-value column):

- X_1 Working capital/Total Assets
- X_5 Sales/Total Assets

The model without insignificant variables is called the **restricted model**, which we will fully re-estimate using logit regression.

Restricted Model Inference

Nobs	4000
PD	1.80%

Initial Parameters	
C	-4.00
Beta2	0
Beta3	0
Beta4	0

	Estimates
C	-2.318
Beta2	-1.420
Beta3	-7.179
Beta4	-1.616

Sum log L
-282.219

Y	1-Y	beta*X	Log L
0	1	-4.61	-0.01
0	1	-4.87	-0.01
0	1	-4.13	-0.02
0	1	-3.43	-0.03
0	1	-4.14	-0.02
0	1	-4.95	-0.01
0	1	-2.50	-0.08
0	1	-2.61	-0.07
0	1	-3.11	-0.04
0	1	-2.93	-0.05
0	1	-3.16	-0.04
0	1	-3.98	-0.02
0	1	-3.43	-0.03
0	1	-2.77	-0.06
0	1	-2.52	-0.08
0	1	-2.64	-0.07
0	1	-2.52	-0.08
0	1	-3.62	-0.03
0	1	-3.65	-0.03
0	1	-4.06	-0.02
0	1	-4.08	-0.02
0	1	-3.71	-0.02
0	1	-4.27	-0.01

Information Matrix

C	Beta2	Beta3	Beta4
62.95	-9.99	2.14	32.30
-9.99	22.99	-0.29	-0.18
2.14	-0.29	0.21	1.18
32.30	-0.18	1.18	27.23

Inverse Info Matrix

C	Beta2	Beta3	Beta4
0.06	0.02	-0.22	-0.06
0.02	0.05	0.00	-0.02
-0.22	0.00	7.43	-0.06
-0.06	-0.02	-0.06	0.11

Inference Table

Parameter	Estimates	Std err	t-stats
C	-2.32	0.24	-9.84
Beta2	-1.42	0.23	-6.21
Beta3	-7.18	2.73	-2.63
Beta4	-1.62	0.32	-4.97

Likelihood Ratio (LR) Test

This simple and practical test examines the difference between the likelihoods of two models.

The special case forms as our restricted model M_0 is nested within the original model M_1.

Null Hypothesis H_0: the restricted model is the 'true' model.

Test statistic is calculated as a ratio of likelihoods:

$$D = -2 \log \frac{f_{\mathbf{Y}}(\widehat{\theta}_0)}{f_{\mathbf{Y}}(\widehat{\theta})} = 2(\widehat{\ell}_{M1} - \widehat{\ell}_{M0})$$

$D \sim \chi^2_k$ with degrees of freedom k equal to the number of restrictions (i.e., removed variables).

Excluding variables that appear insignificant create **a new model**. Compare nested models by their total Likelihoods.

Likelihood Ratio Test	Log L	(Nested Models)	
Model 1 - Unrestricted	-280.53	L	H1: M1 is a significantly different alternative
Model 0 - Restricted	-282.22	L0	H0: M0 is a 'true' model
Deviance, D	3.386994	2(L-L_0)	
p-value (Chi Square)	0.183875	DF=2	two restrictions (two variables removed)

The probability of difference between the models follows Chi Square distri 0.816 and **is not high enough**

We formally **Do Not Reject** H_0 and so the restricted model, M0 is the 'true' model.

If the models are 'no different' in the likelihood they produce, we prefer the compact model.

Credit Ratings Migration

Ratings System

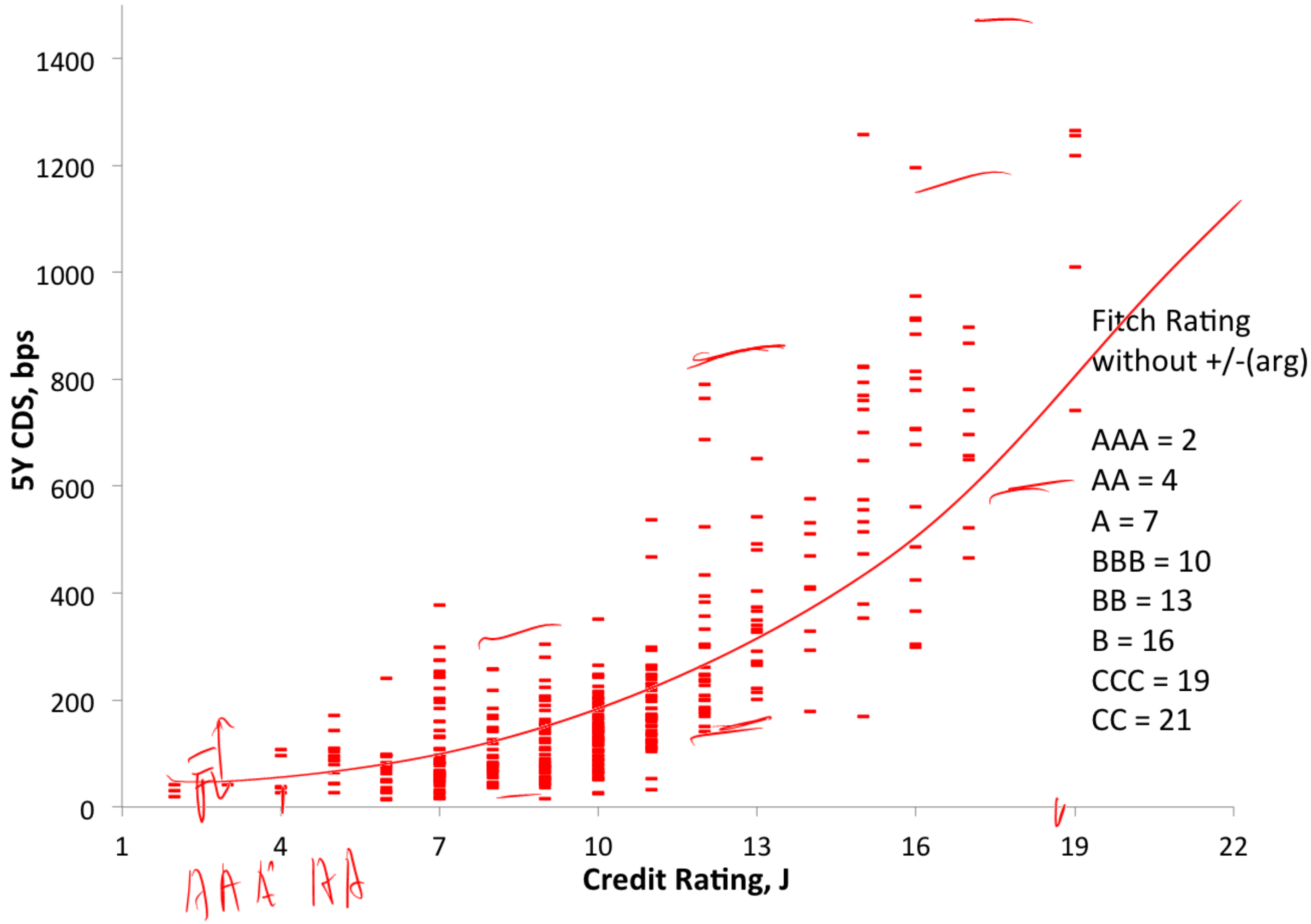
A credit rating system uses a set of grades to rank debt issuers according to their credit quality.

More than 96% of ratings are assigned by Fitch, Moody's and Standard & Poor's that are designated as Nationally Recognized Statistical Rating Organizations (NRSRO) by the U.S. SEC.

- In the past, the rating agencies published tables **matching** a rating with default probability.

- A credit rating mediates the connection between default event and default probability. Quant models infer the PD from credit spreads, while analysts trace changes ΔCDS.

The relationship between ratings and credit spreads (as market indicators) requires a non-linear fit.



Source: **Fitch Ratings**, 2012

Credit Migration

If a bond is re-rated higher then it appreciates. This **rating transition probability** must be reflected in a current price.

$$Z_I = e^{-rT}\left((1-\mathrm{RR})e^{-\lambda T} + \mathrm{RR}\right)$$

To estimate the transition probabilities from fundamentals data the **ordered probit** regression is often applied.

Default event	$\Rightarrow$	Re-rating/Change in spreads
PD	$\Rightarrow$	Rating Transition

Rating Transition Matrix

$$\begin{pmatrix} 1-p & p \\ 0 & 1 \end{pmatrix}$$

S&P Sovereign Transition Matrix

	AAA	AA	A	BBB	BB	B	CCC	CC / D
AAA	0.969	0.031	0.000	0.000	0.000	0.000	0.000	0.000
AA	0.006	0.977	0.011	0.000	0.006	0.000	0.000	0.000
A	0.000	0.030	0.939	0.020	0.001	0.010	0.000	0.000
BBB	0.000	0.000	0.033	0.926	0.024	0.017	0.000	0.000
BB	0.000	0.000	0.001	0.057	0.885	0.056	0.001	0.000
B	0.000	0.000	0.000	0.002	0.063	0.886	0.031	0.018
CCC	0.000	0.000	0.000	0.000	0.001	0.066	0.241	0.693
CC / D	0.000	0.000	0.000	0.000	0.005	0.169	0.003	0.823

Source: Hu et al. 2001. The Estimation of Transition Matrices for Sovereign Credit Ratings. Year 2012 values are similar: the sovereign credit migration is stable.

Latent variable probit

$$\text{Logit } \boldsymbol{Y} \;\Rightarrow\; g(\boldsymbol{X\beta'})^{-1} = p_i \qquad p_i = \Lambda(\boldsymbol{X_i\beta'})$$

$$\text{Probit } \boldsymbol{J} \;\Rightarrow\; \boldsymbol{\beta X(A)} \qquad p_i = \Phi(\boldsymbol{X_i\beta'})$$

The choice of link is the Inverse of Normal CDF $g(p) = \Phi^{-1}(p)$ because it gives $g(\ldots)^{-1} = \Phi(\ldots)$.

For a generalised regression model,

$$\begin{aligned}
\boldsymbol{Y(A)} &= \boldsymbol{\beta X(A)} + \boldsymbol{\epsilon} \\
g(p) &= \boldsymbol{\beta X(A)} \qquad \text{under expectation} \\
p &= \Phi(\boldsymbol{\beta X(A)}).
\end{aligned}$$

Credit Quality Thresholds z_i

Credit risk models assume **a latent variable** A that reflects the creditworthiness of an issuer. $\boldsymbol{A}$ can be estimated by the Merton Model as the normalised firm's value V_0.

There exists a series of thresholds for the latent variable $\boldsymbol{A}$ such that

$$J = \begin{cases} 0 & \text{if} \quad A \leq 0 \\ 1 & \text{if} \quad 0 < A \leq z_1 \\ & \quad \vdots \\ J & \text{if} \quad z_{j-1} < A \end{cases}$$

Credit rating is an observable ordinal variable $j = 0, 1, 2, \ldots, J$

Rating Transition Probability

Once $\hat{\boldsymbol{\beta}}$ are known, the following scheme is used to convert the fundamentals data $\boldsymbol{X_i}(\boldsymbol{A})$ into the rating $\boldsymbol{J_i}$.

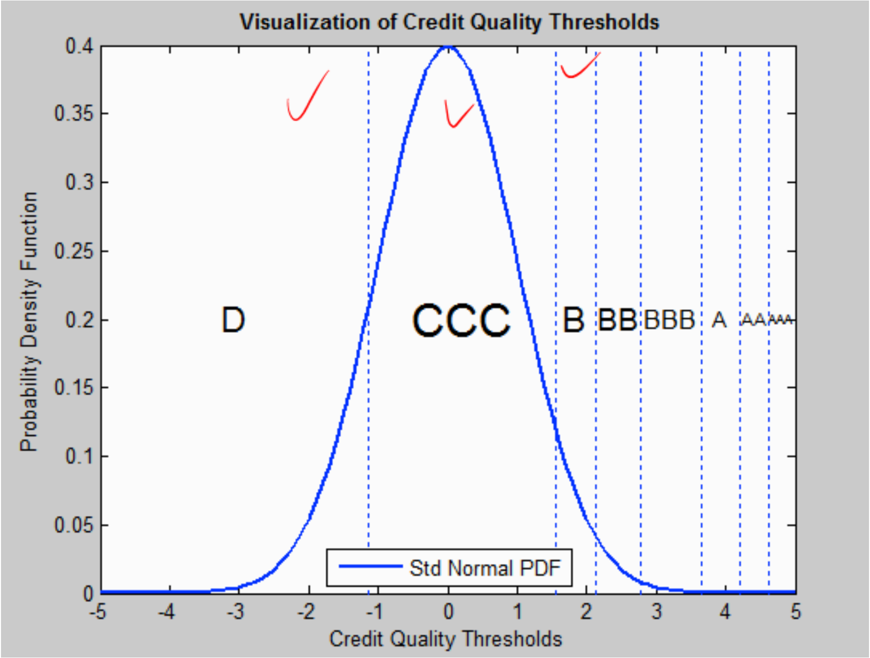
$$\begin{aligned}
\Pr(j=0) &= \Pr(A \leq 0) \\
&= \Pr(\boldsymbol{X\beta'} + \epsilon \leq 0) \qquad \text{where } \epsilon \sim \Phi(0,1) \\
&= \Phi(-\boldsymbol{X\beta'}) \quad \checkmark
\end{aligned}$$

$$\begin{aligned}
\Pr(j=1) &= \Pr(0 < A \leq z_1) \\
&= \Pr(A \leq z_1) - \Pr(A \leq 0) \qquad \text{chunk of prob. mass} \\
&= \Phi(\boldsymbol{X\beta'} + \epsilon \leq z_1) - \Phi(-\boldsymbol{X\beta'}) \\
&= \Phi(z_1 - \boldsymbol{X\beta'}) - \Phi(\boldsymbol{X\beta'}) \quad \checkmark \\
&\ldots
\end{aligned}$$

$$\begin{aligned}
\Pr(j=J) &= \Pr(A > z_{j-1}) \\
&= 1 - \Phi(z_{j-1} - \boldsymbol{X\beta'}) \quad \checkmark
\end{aligned}$$

Ordered Probit (with thresholds)

For a sample of initially CCC debtors, **1.** Credit quality thresholds z_i have to be pre-estimated from frequencies and used in **2.** Calibration of ordered probit – MLE with $p_i = \Phi(\boldsymbol{X_i}\boldsymbol{\beta'})$.



Note: to match the assumption of the firm value $A > 0$, the thresholds are adjusted such that D starts at $z = 0$.

Rating Quantitative Analytics

The task for a rating analyst team is *not limited* to assigning a rating *per se* (which implies a model-dependent PD).

The work extends to **3.** Simulating the past credit history of potential re-rating events for the reference name and **4.** Building a rating transition matrix for that name, sector, etc.

- thresholds z_i are pre-estimated from comparable issuers.
- coefficients $\hat{\beta}$ are calibrated using probit regression model on company fundamentals.

Why is credit migration important?

Basel II Revisions (July 2009) to the *Guidelines for computing capital for incremental risk in the* **Trading Book** note:

- "recent credit market turmoil... losses have not arisen from actual defaults but rather from credit migrations combined with widening of credit spreads and the loss of liquidity."

CVA hedgers rely heavily on CDS (pays on default event, not a re-rating). They are likely to be overpaying.

Recent efforts have been focused on Incremental Risk Charge modelling.

Summary

Please take away the following ideas...

- Default probability (credit risk) is implied in risky bond prices and affects volatility smile.
- Statistical estimation of default probability relies on GLM and is commonly done using a logistic regression.
- Quantitative credit analytics recovers past credit history and builds rating transition matrices.
- If you are interested to estimate probability from data, use logit or probit regression model.
- To choose between full and restricted models apply the likelihood approach (LR test).