Mathematical Models And Statistical Analysis of Credit Risk Management

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE
in
MATHEMATICS AND SCIENCE BASED BUSINESS

Author :	Tianfeng Hou
Supervisor :	Dr. O. van Gaans
2^{nd} corrector :	Prof.dr. R.D. Gill
3^{nd} corrector :	Dr. F.M. Spieksma

Leiden, The Netherlands, September 26, 2014

Mathematical Models And Statistical Analysis of Credit Risk Management

Tianfeng Hou

Mathematical Institute, Universiteit Leiden
P.O. Box 9500, 2300 RA Leiden, The Netherlands

September 26, 2014

Abstract

This thesis concerns mathematical models and statistical analysis of management of default risk for markets, individual obligors, and portfolios. Firstly, we consider to use CPV model to estimate default rate of both Chinese and Dutch credit market. It turns out that our CPV model gives good predictions. Secondly, we study the KMV model, and estimate default risk of both Chinese and Dutch companies based on it. At last, we use two mathematical models to predict the default risk of investors' entire portfolio of loans. In particular we consider the influence of correlations. Our models show that correlation in a portfolio may lead to much higher risks of great losses.

Contents

Chapter 1

Introduction to defaults and losses

Credit risk management is becoming more and more important in today's banking activity. It is the practice of mitigating losses by understanding the adequacy of both a bank's capital and loan loss reserves to any given time. In simple words, the financial engineers in the bank need to create a capital cushion for covering losses arising from defaulted loans. This capital cushion is also called expected loss reserve[2]. It is important for a bank to have good predictions for its expected loss. If a bank keeps reserves that are too high, than it misses profits that could have been made by using the money for other purposes. If the reserve is too low, the bank must unexpectedly sell assets or attract capital, probably leading to a loss or higher costs. Mathematical models are used to predict expected losses. Before we discuss various ways of credit risk modelling we will first look at several definitions.

1.1 How to define the loss

1.1.1 The loss variable

Let us first look at one obligor. By definition, the potential loss of an obligor is defined by a loss random variable

$$\tilde{L} = EAD \times LGD \times L \quad \text{with} \quad L = 1_D, \quad \mathbb{P}(D) = DP,$$

where the exposure at default (EAD) stands for the amount of the loan's exposure in the considered time period, the loss given default (LGD) is a percentage, and stands for the fraction of the investment the bank will lose if default happens. (DP) stands for the default probability. D denotes the

event that the obligor defaults in a certain period of time(most often one year),and $\mathbb{P}(D)$denotes the probability of the event D.
Default rate is the rate at which debt holders default on the amount of money that they owe. It is often used by credit card companies when setting interest rates, but also refers to the rate at which corporations default on their loans. Default rates tend to rise during economic downturns, since investors and businesses see a decline in income and sales while still required to pay off the same amount of debt. So If we invest in debt we want to know or minimize the risk of default.

1.1.2 The expected loss

The expected loss (EL) is the expectation of the loss variable $\tilde{L}$. The definition is

$$EL = \mathbb{E}[\tilde{L}].$$

If EAD and LGD are constants

$$\begin{aligned} EL &= EAD \times LGD \times \mathbb{P}(D) \\ &= EAD \times LGD \times DP. \end{aligned}$$

This formula also holds if EAD and LGD are the expectations of some underlying random variables that are independent of D.

1.1.3 The unexpected losses

Then we turn to portfolio loss. As we discussed before the financial engineers in the bank need to create a capital cushion for covering losses arising from defaulted loans.A cushion at the level of the expected loss will often not cover all the losses. Therefore the bank needs to prepare for covering losses higher than the expected losses, sometimes called the unexpected losses.
A simple measure for unexpected losses is the standard deviation of the loss variable $\tilde{L}$,

$$UL = \sqrt{\mathbb{V}[\tilde{L}]} = \sqrt{\mathbb{V}[EAD \times SEV \times L]}.$$

Here the SEV is the severity of loss which can be considered as a random variable with expectation given by the LGD.

1.1.4 The economic capital

It is not the best way to measure the unexpected loss for the risk capital by the standard deviation of the loss variable, especially if an economic crisis happens. It is very easy that the losses will go far beyond the portfolio's expected loss by just one standard deviation of the portfolio's loss.
It is better to take into account the entire distribution of the portfolio loss. Banks make use of the so-called economic capital.
For instance, if a bank wants to cover 95 percent of the portfolio loss, the economic capital equals the 0.95 th quantile of the distribution of the portfolio loss, where the qth quantile of a random variable $\tilde{L}_{PF}$ is defined as

$$q_\alpha = \inf\{q > 0 \mid \mathbb{P}[\tilde{L}_{PF} \leq q] \geq \alpha\}.$$

The economic capital (EC) is defined as the α - quantile of the portfolio loss $\tilde{L}_{PF}$ minus the expect loss of the portfolio,

$$EC_\alpha = q_\alpha - EL_{PF}.$$

So if the bank wants to cover 95 percent of the portfolio loss, and the level of confidence is set to α =0.95, then the economic capital EC_α can cover unexpected losses in 9,500 out of 10,000 years, if we assume a planning horizon of one year.

Chapter 2

How To Model The Default Probability

2.1 General statistical models

2.1.1 The Bernoulli Model

In statistics, if an experiment only has two future scenarios, A or $\bar{A}$, then we call it a Bernoulli experiment. In our default-only case, every counterparty either defaults or survives. This can be expressed by Bernoulli variable [2],

$$L_i \sim B(1; p_i), \ \ i.e., L_i = \begin{cases} 1 & \text{with probability } p_i, \\ 0 & \text{with probability } 1 - p_i. \end{cases}$$

Next, we assume the loss statistics variables $L_1, ..., L_m$ are independent and regard the loss probabilities as random variables $P = (P_1, ..., P_m) \sim F$ with some distribution function F with support in$[0,1]^m$,

$$L_i \mid P_i = p_i \sim B(1; p_i), \ \ (L_i \mid P = p)_{i=1,...,m} \ \text{ independent.}$$

The joint distribution of the L_i is then determined by the probabilities

$$\mathbb{P}[L_1 = l_1, ..., L_m = l_m] = \int_{[0,1]^m} \prod_{i=1}^n p_i^{l_i} (1 - p_i)^{1-l_i} dF(p_1, ..., p_m),$$

where $l_i \in \{0, 1\}$. The expectation and variance are given by

$$\mathbb{E}[L_i] = \mathbb{E}[P_i], \quad \mathbb{V}[L_i] = \mathbb{E}[P_i](1 - \mathbb{E}[P_i]) \quad (i = 1, ..., m).$$

The covariance between single losses obviously equals

$$\text{Cov}[L_i, L_j] = \mathbb{E}[L_i, Lj] - \mathbb{E}[L_i]\mathbb{E}[L_j] = \text{Cov}[P_i, P_j].$$

The correlation in this model is

$$\text{Corr}[L_i, L_j] = \frac{Cov[P_i, P_j]}{\sqrt{\mathbb{E}[P_i](1-\mathbb{E}[P_i])}\sqrt{\mathbb{E}[P_j](1-\mathbb{E}[P_j])}}.$$

2.1.2 The Poisson Model

There are other models in use than the conditional Bernoulli model of section 2.1.1. For instance, *CreditRisk*$^+$ by Credit Suisse uses a conditional Poisson model[7]. The reason is that *CreditRisk*$^+$ uses generating functions of default probabilities in its calculation rather than the distributions themselves and the generating function of Poisson distributions have a convenient exponential form.
In the Possion model, obligor $i \in \{1, ...m\}$ will default L'_i times in a considered time period, where L'_i is a Poisson random variable with parameter Λ_i, so

$$P\{L'_i = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, \ \ k = 0, 1, 2, ...,$$

where $\lambda = \Lambda_i$ will also be a random variable. So the default vector $(L'_1, ...L'_m)$ consists of Poisson random variables $L'_i \sim \text{Pois}(\Lambda_i)$, where $\Lambda = (\Lambda_1, ..., \Lambda_m)$ is a random vector with some distribution function F with support in$[0, \infty)^m$. Moreover, it is assumed that the conditional random variables $(L'_i \mid \Lambda = \lambda)_{i=1,...,m}$ are independent.
The joint distribution of the L'_i is then determined by the probabilities

$$\mathbb{P}[L'_1 = l'_1, ..., L'_m = l'_m] = \int_{[0,\infty)^m} e^{-(\lambda_1 + ... + \lambda_m)} \prod_{i=1}^m \frac{\lambda_i^{l'_i}}{l'_i!} dF(\lambda_1, ..., \lambda_m),$$

where $l'_i \in \{0, 1, 2, ...\}$. The expectation and variance are given by

$$\mathbb{E}[L'_i] = \mathbb{E}[\Lambda_i], \quad \mathbb{V}[L'_i] = \mathbb{V}[\Lambda_i] + \mathbb{E}[\Lambda_i] \quad (i = 1, ..., m).$$

The covariance satisfies $\text{Cov}[L'_i, L'_j] = \text{Cov}[\Lambda_i, \Lambda_j]$ and the correlation between defaults is

$$\text{Corr}[L'_i, L'_j] = \frac{\text{Cov}[\Lambda_i, \Lambda_j]}{\sqrt{\mathbb{V}[\Lambda_i] + \mathbb{E}[\Lambda_i]}\sqrt{\mathbb{V}[\Lambda_j] + \mathbb{E}[\Lambda_j]}}.$$

It may seem unrealistic that one obligor can default more than once in one time period. However, often the rates Λ_i will be small and then the probability of defaulting more than once will be very small. If we neglect this small probability, this Poisson model becomes the same as the Bernoulli model. More detail can be found in [7].

2.2 The CPV Model and KMV Model

2.2.1 Credit Portfolio View

Credit Portfolio View (CPV)[3] is based upon the argument that default and migration probabilities are not independent of the business cycle. Here we think of all loans being classified in classes of different quality and a migration probability, it take probability that a loan changes from one class to another. In the simplest case there are two classes: in default and not in default. In the latter case the default probability may be viewed as the probability of migrating from 'not in default' to 'in default'. CPV calls any migration matrix observed in a particular year a conditional migration matrix, and the average of conditional migration matrices in a lot of years will give us an unconditional migration matrix. The idea is that the migration probabilities are conditional on the economic situation in that particular year. The economic situation is assumed to be approximately cyclic and therefore its effect is averaged out over a lot of years. During boom times default probabilities run lower than the long term average that is reflected in the unconditional migration matrix; and conversely during recessions default probabilities and downward migration probabilities run higher than the longer term average. This effect is more amplified for speculative grade credits than for investment grade as the latter are more stable even in tougher economic situations.
This adjustment to the migration matrix is done by multiplying the unconditional migration matrix by a factor that reflects the state of the economy. If M be the unconditional transition matrix, then $M_t = (r_t - 1)A + M$ is the Conditional transition matrix. How do we derive the factor r_t?
Here $A = a_{ij}$ is a suitable matrix such as $a_{ij} \geq 0$ for $i < j$ and $a_{ij} \leq 0$ for $i > j$. The factor r_t is just chosen to be the conditional probability of default in period t divided by the unconditional (or historical) probability of default. This is expressed as follows:

$$r_t = \frac{P_t}{\bar{P}},$$

where P_t is the conditional probability of default in period t, and $\bar{P}$ is the unconditional probability.
Now P_t itself is modelled as a logistic function of an index value Y_t,

$$\frac{1}{1+\exp(-Y_t)}.$$

The index Y_t is derived using a multi-factor regression model that considers a number of macro economic factors,

$$Y_t = \beta_0 + \sum_{k=1}^{K} \beta_k X_{k,t} + \varepsilon_t,$$

where $X_{k,t}$ are the macroeconomics factors at time t, w_k are coefficients of the corresponding macroeconomics factors, w_0 is the intercept of the linear model, and ε_t is the residual random fluctuation of Y_t

2.2.2 The KMV-Model

The KMV-Model is a well-known industry model [1].This model was created by the United States KMV corporation and it is named by three founders of this company, Kealhofer, MeQuow, and Vasicek. The idea of the KMV model is based on whether the firm's asset values will fall below a certain critical threshold or not. Let A_t^i denote the asset value of firm $i \in \{1,,m\}$. If after a period of time T the firm's asset value A_T^i is below this threshold C_i then we say the firm is in default. Otherwise the firm survived the considered time period. We can represent this model in a Bernoulli type model. Indeed, consider the random variable L_i defined by

$$L_i = 1_{\{A_T^{(i)} < C_i\}}.$$

This random variable has a Bernoulli distribution,

$$B(1;\mathbb{P}[A_T^{(i)} < C_i]) \ \ (i = 1,...,m).$$

The classic Black-Scholes-Merton model [10] gives a model for the firm's asset value.

$$A(t) = Cexp(\alpha t + \theta W(t)),$$

where $C > 0$ is constant, α, θ are constant and W is a Brownian motion.
The logarithmic return over time T is then:

$$\begin{aligned} \ln A(T) - \ln A(0) &= \ln C + \alpha T + \theta W(T) - (\ln C + 0) \\ &= \alpha T + \theta W(T). \end{aligned}$$

Note that $\theta W(T) \sim N(0,T)$
The term αT is deterministic and can be absorbed in the threshold, so without loss of generality we can take $\alpha = 0$. Further, we will think of the random part as consisting if two separate parts: one determined by the economic situation and one being specific for the individual obligor. Thus we arrive at the following formula for the (logarithmic) asset return at time T:

$$r_i = \beta_i \phi_i + \varepsilon_i \quad (i = 1, ...m).$$

Here, ϕ_i is called the composite factor of firm i which is a standard normally distributed random variable describing the state of the economic environment of the firm. β_i is the sensitivity coefficient, which captures the linear correlation of r_i and ϕ_i. The normal random variable ε_i stands for the residual part of r_i, it means that the return r_i differs from the prediction $\beta_i \phi_i$ based on the economic situation by an error ε_i, which is called the idiosyncratic part of the return.
We rescale the (logarithmic) asset value return to become a standard normal random variable,

$$\widetilde{r}_i = \frac{r_i - \mathbb{E}[r_i]}{\mathbb{V}[r_i]} \quad (i = 1, ...m).$$

With the coefficient R_i defined by

$$R_i^2 = \frac{\beta_i^2 \mathbb{V}[\phi_i]}{\mathbb{V}[r_i]} \quad (i = 1, ...m),$$

and with the same sign as β_i we get a representation

$$r_i = R_i \phi_i + \varepsilon_i \quad (i = 1, ...m).$$

Here R_i is given above, ϕ_i means the company's composite factor, and ε_i is the idiosyncratic part of the company's asset value log-return.
Observe that

$$r_i \sim N(0,1), \;\; \Phi_i \sim N(0,1), \;\; \text{and} \;\; \varepsilon_i \sim N(0, 1 - {R_i}^2).$$

As in the Bernoulli Model, the joint distribution of the L_i is then determined by the probabilities

$$\mathbb{P}[L_1 = l_1, ..., L_m = l_m] = \int_{[0,1]^m} \prod_{i=1}^{n} p_i^{l_i} (1 - p_i)^{1-l_i} dF(p_1, ..., p_m).$$

Here what we should get clear is the distribution function F which is still a degree of freedom in the model. The event of default of firm i at time T corresponds to $r_i < c_i$. This is equivalent to

$$\varepsilon_i < c_i - R_i\phi_i.$$

Denoting the one-year default probability of obligor i by $\widetilde{p_i}$, we have $\widetilde{p_i} = \mathbb{P}[r_i < c_i]$. As $r_i \sim N(0,1)$, we get

$$c_i = N^{-1}[\widetilde{p_i}] \quad (i = 1, ...m).$$

Here $N[\centerdot]$ denotes the CDF (cumulative distribution function) of the standard normal distribution. We can easily replace ε_i by a standardized normal random variable $\widetilde{\varepsilon_i}$ by means of

$$\widetilde{\varepsilon_i} < \frac{N^{-1}[\widetilde{p_i}] - R_i\Phi_i}{\sqrt{1 - R_i{}^2}}, \ \ \widetilde{\varepsilon_i} \sim N(0,1).$$

Because of $\widetilde{\varepsilon_i} \sim N(0,1)$, the one-year default probability of obligor i conditional on the factor Φ_i can be represented

$$\widetilde{p_i}(\phi_i) = N[\frac{N^{-1}[\widetilde{p_i}] - R_i\phi_i}{\sqrt{1 - R_i{}^2}}] \quad (i = 1, ...m),$$

Finally, if we assume that the distribution function F is that of a multivariate normal distribution, then we can express it as

$$F(p_1, ..., p_m) = N_m[p_1{}^{-1}(\widetilde{p_1}), ..., p_m{}^{-1}(\widetilde{p_m}); \Gamma],$$

where $N_m[\centerdot; \Gamma]$ denotes the cumulative centered Gaussian distribution with correlation matrix Γ, and Γ means the asset correlation matrix of the log-returns r_i.

In the computations above, we have assumed that firm i is in default at time T precisely when its asset value at time T is below a certain threshold. If T is the maturity time of the debt, it is more realistic to assume that firm i is in default at time T if at some moment t between 0 and T its asset value has been below the threshold. In that case, one can use the theory of option pricing for the classic Black-Scholes-Merton mode, as is briefly reviewed next.

The process of KMV model

The process of KMV model can be divided into four steps.

The first step is: Estimate the company's asset value and its volatility.
In 1973 Fisher Black and Myron Scholes found the first solution for the valuation of options called Black-Scholes pricing model [10]. In 1974 Merton implemented this option pricing model into a bond pricing model [10].
In Merton's model, the option is maturing in τ periods. The firm's asset

value V satisfies the following,

$$E = V \times N(d_1) - B \times e^{-rt} \times N(d_2),$$
$$d_1 = \frac{ln(\frac{V}{B}) + (r + \frac{1}{2} \times \sigma_v^2)\tau}{\sigma_v \sqrt{(\tau)}},$$
$$d_2 = d_1 - \sigma_v \sqrt{(\tau)}.$$

Here E is the market value of the firm, V is the asset value of the company, B is the price for the loan, r is the interest rate, σ_v is the volatility of asset value. τ is put option expiration date or in the case of a bond, the maturing time. $N(d)$ is the Cumulative standard normal distribution probability function.
Moreover two founders of the KMV corporation, Oldrich Vasicek and Stephen Kealhofer extended Merton's model by relating the volatility of the firm's market value to the volatility of its asset value [10].

$$\sigma_s = (\frac{V \times N(d_1) \times \sigma_v}{E}).$$

Here σ_s is the volatility of firm's market value.
In short, we have in general form:

$$\hat{E} = f(V, \hat{B}, \hat{r}, \sigma_v, \hat{\tau}),$$

and,

$$\hat{\sigma_s} = g(\sigma_v).$$

Since we have two equations and two unknowns(V, σ_v), $\hat{\sigma_s}$ is the volatility of market value. We use a standard iterative method to find V and σ_v.

The second step: Find the default point.
The default happens when the value of the firm falls below"default point". According to the studies of the KMV, some of the companies will not default while their firm's asset reach the level of total liabilities due to the different debt structure. Thus DPT is somewhere between total liabilities and current liabilities, as below:

$$DPT = SL + \alpha LL, 0 \leqslant \alpha \leqslant 1.$$

Under a large empirical investigation, KMV found that a good choice of Default Point is to take it equal to the short-term liabilities plus half of long-term liabilities[11],

$$DPT = SL + 0.5LL.$$

Here DPT is the default point, SL is the short-term liabilities, LL is the long-term liabilities.

The Third step: find the default-distance (DD).

The default-distance (DD) is the number of standard deviations between the mean of asset value's distribution and the default point. After we get the implied V, σ_v and the default point, the default-distance DD can be computed as follows:

$$DD = \frac{E(V) - DP}{E(V) \times \sigma_v}.$$

The Fourth step: Estimate the company's expected default probability (EDF)

The Expected default probability (EDF) is determined by mapping the default distance (DD) with the expected default frequency.

As the firm's asset value of Merton model is normally distributed, the expectation $E(V)$ of V is $V_0 \exp(ut)$, which is log-normally distributed. Thus the DD expressed in units of asset return standard deviations at the time horizon T is

$$DD = \frac{\ln(\frac{V_{A0}}{DPT_T}) + (\mu - 0.5\sigma_A^2)T}{\sigma_v \sqrt{T}}.$$

Here V_{A0} is the current market value of the assets, DPT_T is the default point at time horizon T, μ is the expected annual return on the firm's assets,σ_A is the annualized asset volatility.

So the corresponding theoretical implied default frequency (EDF) at one year interval is

$$EDF_{Theoretical} = N(-\frac{\ln(\frac{V_{A0}}{DPT_T}) + (\mu - 0.5\sigma_A^2)T}{\sigma_v \sqrt{T}}) = N(-DD).$$

The asset value is not exactly normally distributed in practice. Based on the one-to-one mapping relations between the default distance DD and the expected default frequency(EDF), the length of the distance to a certain extent reflects the company's credit status, and thus evaluates the level of competitiveness of the enterprise.

Chapter 3

Use CPV model to estimate default rate of Chinese and Dutch credit market

In this chapter we want to use the CPV model as described in Section 2.2.1 to estimate the default rate(DR) of Chinese and of the Dutch credit market. We will use real world data of the Chinese joint-equity commercial bank and the Dutch national bank.

3.1 Use CPV model to estimate default rate of Chinese credit market

3.1.1 Macroeconomic factors and data

In the CPV model macroeconomic factors drive the default rate. Typical candidates for macroeconomic factors are Consumer Price Index(CPI), financial expenditure(FE), urban disposable incomes(DI), Business Climate Index(BSI), interest rate(APR), Gross Domestic Product(GDP) and other variables reflecting the macroeconomy of a country.

In our case study we choose Consumer Price Index(CPI), unemployment rate(UR), financial expenditure(FE), urban disposable incomes(DI), Fixed asset investment price index(FAIPI), money supply(MS), Business Climate Index(BSI),interest rate(APR), Gross Domestic Product(GDP), and the growth rate of GDP(Growth) to be the macroeconomic factors.

Let us briefly summaries the meaning of these quantities.

A consumer price index (CPI) measures changes in the price level of a market basket of consumer goods and services purchased by households. The annual percentage change in a CPI is used as a measure of inflation. In most countries, the CPI is one of the most closely watched national economic statistics.

Unemployment (or joblessness) occurs when people are without work and actively seeking work. The unemployment rate (UR) is a measure of the prevalence of unemployment and it is calculated as a percentage by dividing the number of unemployed individuals by all individuals currently in the labor force. During periods of recession, an economy usually experiences a relatively high unemployment rate.

In National Income Accounting, government spending, financial expenditure (FE), or government spending on goods and services includes all government consumption and investment but excludes transfer payments made by a state. It can reflect the strength of the government finance and the future direction of the national economy.

Disposable income (DI) is total personal income minus personal current taxes.

Fixed asset investment price index (FAIPI) reflects the trend and degree of changes in prices of investment in fixed assets. It is calculated as the weighted arithmetic mean of the price indices of the three components of investment in fixed assets (the investment in construction and installation, the investment in purchases of equipment and instrument and the investment in other items).

Money supply (MS) is the total amount of monetary assets available in an economy at a specific time.

Business climate index (BSI) is the index of general economic environment comprising of the attitude of the government and lending institutions toward businesses and business activity, attitude of labor unions toward employers, current taxation regimen, inflation rate, and such.

Interest rate is the rate at which interest is paid by a borrower (debtor) for the use of money that they borrow from a lender (creditor).

Gross domestic product is defined by OECD as ”an aggregate measure of production equal to the sum of the gross values added of all resident institutional units engaged in production (plus any taxes, and minus any subsidies, on products not included in the value of their outputs.

MY Preliminary data
We use time series data on a quarter base over the years 2009-2013. The Chinese joint-equity commercial bank does not have a united definition of default. Instead it use five-category assets classification for the main method for risk management. Comparing the definition of the probability of the non-performing loan in five-category assets classification and the default rate, they are similar. So we choose the probability of the non-performing loan to be the default rate. The data of probability of the non-performing loan is from the official website of China Banking Regulatory Commission. [16].
The data of all the macroeconomic factors is from the official website of National Bureau Of Statistics Of China. [15].

Table 3.1: *All of the required data*

DR	CPI	GDP	Growth	UR	FE	DI	FAIPI	APR	MS	BSI
1.17%	100.03	69816.92	6.6%	4.3%	12810.90	4833.90	98.80	2.3%	502156.67	105.60
1.03%	99.06	78386.68	7.5%	4.3%	16091.70	4022.00	96.10	2.3%	552553.64	115.90
0.99%	98.83	83099.73	8.2%	4.3%	16300.20	4117.40	96.40	2.3%	578402.38	124.40
0.95%	99.10	109599.48	9.2%	4.3%	31097.13	4201.40	99.00	2.3%	597157.51	130.60
0.86%	101.90	82613.39	12.1%	4.1%	14330.00	5308.00	101.90	2.3%	637209.67	132.90
0.80%	102.50	92265.44	11.2%	4.1%	19481.40	4449.10	103.60	2.3%	652611.44	135.90
0.76%	102.80	97747.91	10.7%	4.1%	20693.60	4576.70	103.50	2.5%	686009.97	137.90
0.70%	103.16	128886.06	10.4%	4.1%	35070.00	4775.60	105.40	2.5%	711989.19	138.00
0.70%	104.93	97479.54	9.8%	4.1%	18053.60	5962.80	106.50	3.0%	742715.52	140.30
0.60%	105.23	109008.57	9.7%	4.1%	26381.50	5078.70	106.70	2.9%	767204.88	137.90
0.60%	105.60	115856.56	9.5%	4.1%	25045.50	5259.40	107.30	3.5%	780394.05	135.60
0.60%	105.50	150759.38	9.3%	4.1%	39521.40	5508.90	105.70	3.3%	831304.67	127.80
0.63%	104.07	108471.97	7.9%	4.1%	24118.10	6796.30	102.30	3.3%	872878.60	127.30
0.65%	103.50	119531.12	7.7%	4.1%	29774.90	5712.20	101.60	3.3%	904881.34	126.90
0.70%	102.93	125738.46	7.6%	4.1%	30226.30	5918.10	100.20	3.0%	929218.58	122.80
0.72%	102.67	165728.55	7.7%	4.1%	41592.70	6138.10	100.30	3.0%	951798.71	124.40
0.77%	102.33	118862.08	7.7%	4.1%	27036.70	7427.30	100.20	3.0%	1008862.82	125.60
0.80%	102.40	129162.37	7.6%	4.1%	32677.30	6221.80	99.90	3.0%	1043041.58	120.60
0.83%	102.47	139075.79	7.7%	4.1%	31818.30	6519.90	100.10	3.0%	1063615.98	121.50
0.86%	102.60	181744.97	7.7%	4.1%	48211.70	6786.00	100.90	3.0%	1085336.12	119.50

Data adjusted by CPI Index and after seasonal adjustment
In the data table above, financial expenditure, urban disposable incomes, money supply, Gross Domestic Product(GDP), will influenced by the CPI Index. So If we want to analysis these data, we will calculate the CPI Index first, and adjusted these factors by it.
For calculating the CPI Index,we use the CPI of 1 quarter 2009 as base.(that is, the CPI Index of 1 quarter 2009 is 1). We obtain

$$CPI_I n = CPI_n \times CPI_{n-1} \times ... \times CPI_{base}.$$

After the data adjusted by CPI Index, we found that several macroeconomic factors such as financial expenditure, urban disposable incomes, fixed asset investment price index, gross domestic product(GDP), have strong seasonal component. So we will use seasonal adjustment for removing them. In our case study, we use Eviews 6, seasonal Adjustment, X12 method [14] to adjust the data.
Then as the CPV model relates the default probability P_t to an index Y_t by $P_t = \frac{1}{1+e^{-Y_t}}$,we can get Y_t for every quarter. The results in the table below.

Table 3.2: *Data adjusted by CPI Index and after seasonal adjustment*

DR	Y	CPI	Index	GDP	Growth	UR	FE	DI	FAIPI	APR	MS	BSI
1.17%	-4.4364	1.0003	1	86349.97	6.6%	4.3%	12810.9	4203.28	98.8	2.25%	501697	105.6
1.03%	-4.56526	0.9906	0.99	84377.67	7.5%	4.3%	16254.2	4314.364	96.1	2.25%	554012	115.9
0.99%	-4.60527	0.9883	0.98	83932.6	8.2%	4.3%	16632.9	4432.455	96.4	2.25%	593866.9	124.4
0.95%	-4.64692	0.991	0.97	88650.05	9.2%	4.3%	32058.9	4509.752	99	2.25%	617139.6	130.6
0.86%	-4.74736	1.019	0.99	103787.6	12.1%	4.1%	14474.7	4662.155	101.9	2.25%	642757.2	132.9
0.80%	-4.82028	1.025	1.01	96279.8	11.2%	4.1%	19288.5	4678.001	103.6	2.25%	641601.8	135.9
0.76%	-4.87198	1.028	1.04	94658.52	10.7%	4.1%	19897.7	4642.651	103.5	2.50%	663523.6	137.9
0.70%	-4.95482	1.0316	1.075	98248.59	10.4%	4.1%	32623.3	4625.407	105.4	2.50%	664308.8	138
0.70%	-4.95482	1.0493	1.13	108100.5	9.8%	4.1%	15976.6	4588.409	106.5	3.00%	655794.3	140.3
0.60%	-5.10998	1.0523	1.19	107107.3	9.7%	4.1%	22169.3	4532.268	106.7	2.85%	640626	137.9
0.60%	-5.10998	1.056	1.25	107543.7	9.5%	4.1%	20036.4	4438.88	107.3	3.50%	627696.3	135.6
0.60%	-5.10998	1.055	1.32	107297.1	9.3%	4.1%	29940.5	4345.317	105.7	3.25%	632086.4	127.8
0.63%	-5.06089	1.0407	1.38	98874.12	7.9%	4.1%	17476.9	4282.374	102.3	3.25%	630482.1	127.3
0.65%	-5.02943	1.035	1.42	111357.5	7.7%	4.1%	20968.2	4271.938	101.6	3.25%	633623.9	126.9
0.70%	-4.95482	1.0293	1.47	115196.6	7.6%	4.1%	20562.1	4247.294	100.2	3.00%	635568.9	122.8
0.72%	-4.92645	1.0267	1.5	116462.1	7.7%	4.1%	27728.5	4260.626	100.3	3.00%	636913	124.4
0.77%	-4.85881	1.0233	1.54	97544.06	7.7%	4.1%	17556.3	4193.732	100.2	3.00%	652641.5	125.6
0.80%	-4.82028	1.024	1.58	114382	7.6%	4.1%	20681.8	4181.852	99.9	3.00%	656637	120.6
0.83%	-4.78317	1.0247	1.62	124180.1	7.7%	4.1%	19640.9	4245.933	100.1	3.00%	660139.6	121.5
0.86%	-4.74736	1.026	1.66	123550.7	7.7%	4.1%	29043.2	4256.337	100.9	3.00%	656301.1	119.5

3.1.2 Model building

In CPV model,Y_t is an index value derived using a multi-factor regression model[5] that considers a number of macro economic factors, where t denotes the time period,

$$Y_t = \beta_0 + \sum_{k=1}^{K} \beta_k X_{t_k} + \varepsilon_t.$$

So in our case study

$$Y_t = \beta_0 + \beta_1 CPI + \beta_2 GDP + \beta_3 Growth + \beta_4 UR + \beta_5 FE + \beta_6 DI + \beta_7 FAIPI + \beta_8 APR + \beta_9 MS + \beta_{10} BSI.$$

Table 3.3: *The regression results*

	Coefficients	Std. Error	t value	$Pr(>\mid t \mid)$
Intercept	3.52E-01	4.09E+00	0.086	0.93324
data1$CPI	-1.37E+01	3.27E+00	-4.184	0.00236
data1$GDP	1.79E-06	1.67E-06	1.072	0.31177
data1$Growth	-1.10E-02	2.22E-02	-0.496	0.63154
data1$UR	5.34E+01	4.65E+01	1.149	0.28021
data1$FE	-8.73E-06	1.88E-06	-4.657	0.00119
data1$DI	3.62E-04	3.62E-04	1.001	0.34319
data1$FAIPI	6.27E-02	1.44E-02	4.348	0.00186
data1$APR	2.58E-01	9.24E+00	0.028	0.97834
data1$MS	2.29E-06	8.87E-07	2.580	0.0297
data1$BSI	-2.14E-02	5.68E-03	-3.770	0.00442
Multiple R-squared	9.84E-01	*Adjusted R-squared*	9.67E-01	
F-statistic	56.53	p-value	6.81E-07	
Residual standard error	0.03488			

In this regression results table above, the R-squared is 0.984, Adjusted R-squared is 0.967, F-statistic is 56.53. P-value is 6.81×10^{-7}. This means that the hypothesis H_0 :" all regression coefficients zero" is strongly rejected, so there is explanatory power in this model. But in several individual t-tests the p-values are large. One reason may be multi-collinearity, t-test measures effect of a regressor, partial to all other regressors. Due to correlation between regressor, an individual regressors is not contributing a lot of extra information. The other reason may be that some of the regressors do not influence the default rate at all.
The method we will use next are The backward elimination procedure and incremental F-test for selecting the regressors.

Backward elimination procedure

Table 3.4: *backward elimination procedure table*

Start AIC=-128.2	Df	Sum of Sq	RSS	AIC
APR	1	0.0000009	0.01095	-130.2
Growth	1	0.0002997	0.011249	-129.66
none	1		0.010949	-128.2
DI	1	0.0012179	0.012167	-128.09
GDP	1	0.0013972	0.012347	-127.8
UR	1	0.0016059	0.012555	-127.47
MS	1	0.0080977	0.019047	-119.13
BSI	1	0.0172868	0.028236	-111.26
CPI	1	0.0213017	0.032251	-108.6
FAIPI	1	0.0229941	0.033944	-107.58
FE	1	0.0263803	0.03733	-105.67
Step:AIC=-130.2				
	Df	Sum of Sq	RSS	AIC
Growth	1	0.0003048	0.011255	-131.65
none			0.01095	-130.2
DI	1	0.0016712	0.012622	-129.36
GDP	1	0.0019104	0.012861	-128.99
UR	1	0.0020643	0.013015	-128.75
MS	1	0.0081016	0.019052	-121.13
BSI	1	0.0234519	0.034402	-109.31
CPI	1	0.0239093	0.03486	-109.04
FAIPI	1	0.0266847	0.037635	-107.51
FE	1	0.0275626	0.038513	-107.05
Step:AIC=-131.65				
	Df	Sum of Sq	RSS	AIC
none			0.011255	-131.65
DI	1	0.0020819	0.013337	-130.26
GDP	1	0.0023044	0.01356	-129.93
UR	1	0.0025917	0.013847	-129.51
MS	1	0.0078706	0.019126	-123.05
BSI	1	0.0235256	0.034781	-111.09
CPI	1	0.0239376	0.035193	-110.85
FAIPI	1	0.0270197	0.038275	-109.17
FE	1	0.0278528	0.039108	-108.74

The AIC is used for backward elimination. AIC = 2 log(likelihood) + 2p with p the number of parameters in the model. Smaller values point to better fitting models. Each variable is removed from the model in turn, and the resulting AIC's are reported. For eight regressors the AIC deteriorates (becomes larger) by removal, so these variables are important. For two regressors removal makes the AIC smaller (better), so these regressors are candidates for removal. After we remove them the model is $Y_t = \beta_0 + \beta_1 CPI + \beta_2 GDP + \beta_3 UR + \beta_4 FE + \beta_5 DI + \beta_6 FAIPI + \beta_7 MS + \beta_8 BSI$ The regression results are in the table below.

In this regression results table 3.5, the R-squared is 0.984, Adjusted R-squared is 0.9722, F-statistic is 83.99. P-value is 9.12×10^{-9}. This also means H_0 :" all regression coefficients are zero" is strongly rejected, so there is explanatory power in this model . But for the individual t-tests, the p-values of Gross Domestic Product (GDP), unemployment rate (UR), and urban disposable income (DI), are still big.

Incremental F-test
After the Backward elimination procedure, We will use incremental F-test

Table 3.5: *regression results after removing Growth and APR*

	Coefficients	Std. Error	t value	$Pr(>\mid t \mid)$
(Intercept)	3.96E-01	3.542e+00	0.112	0.91297
data1$CPI	-1.37E+01	2.622e+00	-5.217	0.000287
data1$GDP	1.97E-06	1.378e-06	1.426	0.18151
data1$UR	6.01E+01	3.778e+01	1.592	0.139804
data1$FE	-8.78E-06	1.709e-06	-5.139	0.000324
data1$DI	2.55E-04	1.701e-04	1.501	0.161572
data1$FAIPI	6.28E-02	1.309e-02	4.795	0.000558
data1$MS	2.24E-06	8.092e-07	2.773	0.018114
data1$BSI	-2.08E-02	4.297e-03	-4.837	0.000522
Multiple R-squared:	9.84E-01	Adjusted R-squared	0.9722	
F-statistic:	83.99	p-value	9.12E-09	

to null test hypotheses, comparing Full and Reduced Models. We fit a series of models and construct the F-test, using the Anova function from the car package (type II SS).

Table 3.6: *Anova Table (Type II tests)*

Response: data1$Y				
	Sum Sq	Df	F value	$Pr(>\mid F \mid)$
data1$CPI	0.0278528	1	27.2211	0.0002867
data1$GDP	0.0020819	1	2.0346	0.1815098
data1$UR	0.0025917	1	2.5329	0.139804
data1$FE	0.0270197	1	26.4069	0.0003239
data1$DI	0.0023044	1	2.2522	0.1615722
data1$FAIPI	0.0235256	1	22.9921	0.0005578
data1$MS	0.0078706	1	7.6921	0.0181142
data1$BSI	0.0239376	1	23.3948	0.0005216
Multiple R-squared:	0.9839	Adjusted R-squared	0.9722	
Residuals	0.0112553	11		

In the anova table 3.6 we can also see the p-values of F-tests: Gross Domestic Product (GDP), unemployment rate (UR), and urban disposable income (DI), are large. So in the final we will remove these regressors. The final model is

$$Y_t = \beta_0 + \beta_1 CPI + \beta_2 FE + \beta_3 FAIPI + \beta_4 MS + \beta_5 BSI$$

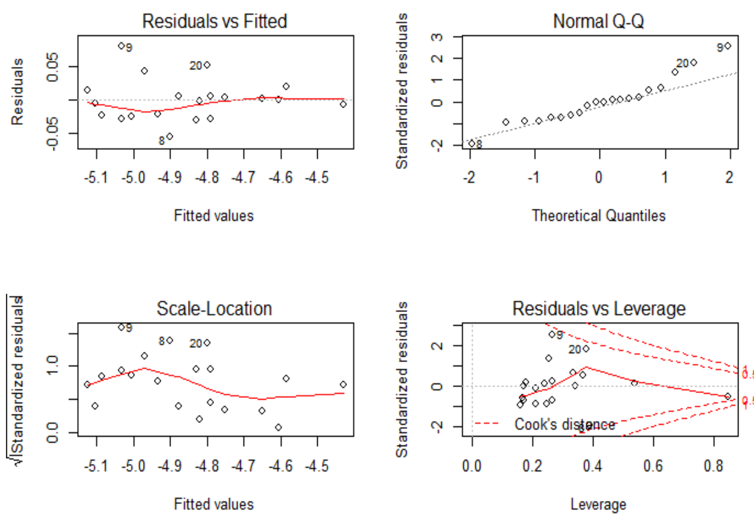
According to the table of The regression results of final model, we can get the model of Y_t,

Table 3.7: *The regression results of final model*

	Coefficients	Std. Error	t value	$Pr(>\|t\|)$
(Intercept)	5.99E+00	5.71E-01	10.491	5.14E-08
data1$CPI	-1.68E+01	1.40E+00	-11.978	9.58E-09
data1$FE	-8.20E-06	1.67E-06	-4.898	0.000235
data1$FAIPI	7.50E-02	1.06E-02	7.083	5.48E-06
data1$MS	1.83E-06	4.26E-07	4.298	0.000737
data1$BSI	-1.78E-02	2.49E-03	-7.156	4.89E-06
Multiple R-squared:	0.9733		Adjusted R-squared	0.9637
F-statistic	102	p-value	1.67E-10	

$$Y_t = 5.99 - 16.8CPI - 0.0000082FE + 0.075FAIPI + 0.00000183MS - 0.0178BSI$$

Diagnostics



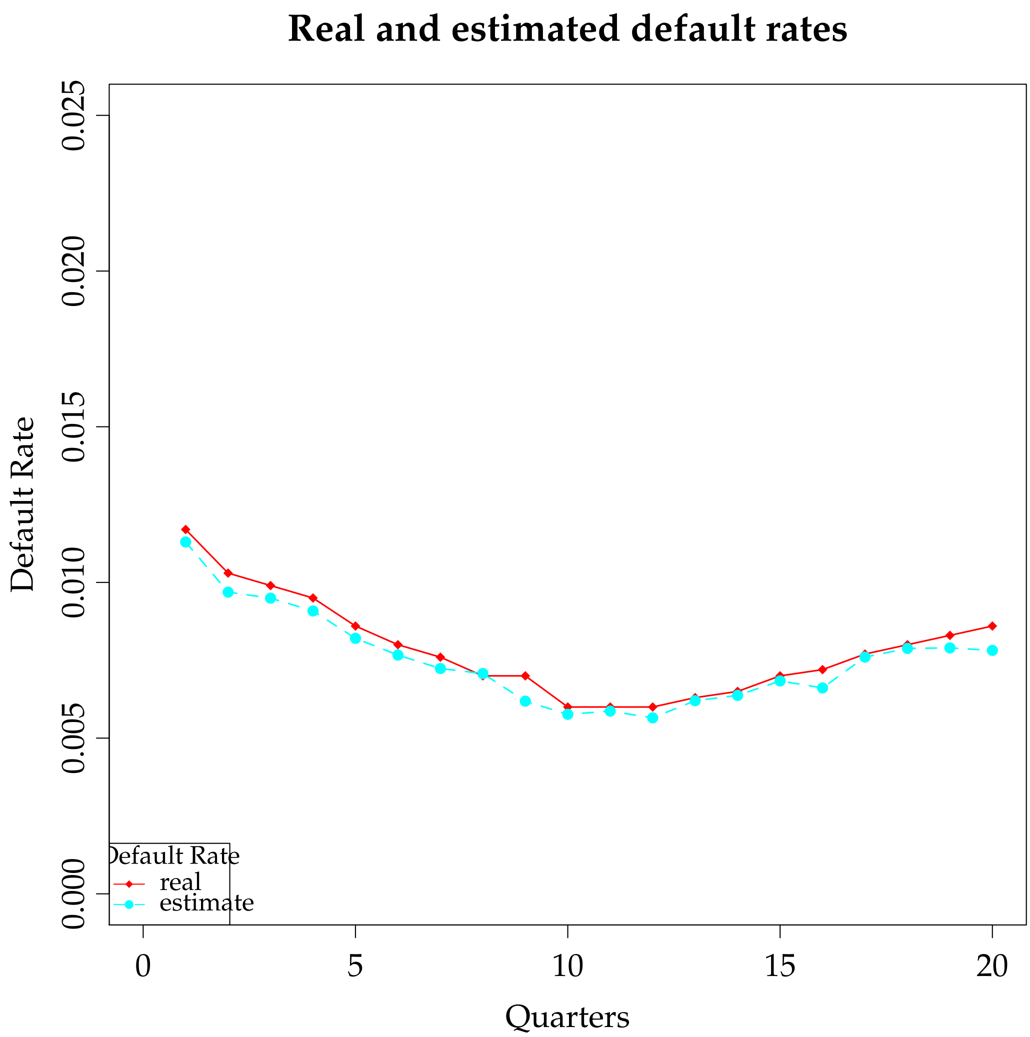
Plot residuals vs fitted values is used for checking constant variance. There are no indications that variance increases with mean.
Normal QQ-plot is used for checking normality. Points lay reasonably well on a straight line, no indications of deviations from normality.
Plot of leverage vs standardized residuals can be used to check for potential influence (leverage), and regression outliers. There are 3 observations with leverage exceeding the threshold 2*p/n=2*5/20=0.5. The standardized residuals are not large for these observations, though, So it does not look problematic.
It may be conclude that the curve fits the data fairly well.

3.1.3 Calculating the default rate

The formula for Y_t with the coefficients fitted to the data as derived in Section 3.12 can be used to compute the default probabilities. Table 3.8 lists the real default probabilities and those computed by means of the formula for Y_t. Below these numbers are shown in a picture.

Table 3.8: *comparing with real default rate and estimate default rate*

Real Default Rate	Estimate Default Rate
0.0117	0.011299
0.0103	0.009689
0.0099	0.009496
0.0095	0.009085
0.0086	0.008205
0.008	0.007668
0.0076	0.007236
0.007	0.007075
0.007	0.006188
0.006	0.005765
0.006	0.005867
0.006	0.005652
0.0063	0.006202
0.0065	0.006373
0.007	0.006837
0.0072	0.006612
0.0077	0.007602
0.008	0.007881
0.0083	0.007898
0.0086	0.007818



3.1.4 Conclusion and discussion

The estimation of the parameters of the model yields a formula that fits the data quite well. From this point of view the model seems good.
Once the parameters of the model have been fitted to the data, the model can be used to make prediction. In order to evaluate the prediction quality of the model, we use it to predict the default rate of the 20th quarter and compare it with the trivial ”tomorrow is same as today” prediction.

According to the table of The regression results,we can get the model of Y_t based on the first 19 quarters.

Table 3.9: *The regression results according to the first 19 quarters*

	Coefficients	Std. Error	t value	$Pr(>\mid t \mid)$
(Intercept)	5.778e+00	5.286e-01	10.932	6.34e-08
data1$CPI	-1.599e+01	1.327e+00	-12.045	2.00e-08
data1$FE	-8.692e-06	1.539e-06	-5.647	7.97e-05
data1$FAIPI	6.852e-02	1.015e-02	6.751	1.36e-05
data1$MS	1.464e-06	4.277e-07	3.423	0.00454
data1$BSI	-1.531e-02	2.578e-03	-5.936	4.94e-05
Multiple R-squared:	0.9792		Adjusted R-squared	0.9712
F-statistic	122.3	p-value	1.851e-10	

$$Y_t = 5.778 - 15.99CPI - 0.000008692FE + 0.06852FAIPI + \\ 0.000001464MS - 0.01531BSI$$

We put the macroeconomic historic data of the 20th quarter into the model above we can easily get the estimated default rate of the 20th quarter is 0.007882. Comparing this estimated default rate with the real default 0.0086 we can see the difference is not big. However if we compare the estimated default rate 0.007882 with the real default rate of the 19th quarter, which is 0.0083, we can find the real default rate of the 19th quarter is much closer to the real default rate of the 20th quarter. Hence the "tomorrow is same as today" prediction is better.

We see that the prediction of the default rate in the 20th quarter made by the model is not bad at all. However, it is not possible to conclude that it is better than prediction made by much simpler models. A more thorough evaluation of the model would require more predictions and comparison of them with the real rates. A test of the model by using 10 data points to fit the coefficients and using the other 10 data points to evaluate the predictions was not successful. There are too many parameters to fit by just 10 data points. A thorough test of the model would require more data.

3.2 Use CPV model to estimate default rate of Dutch credit market

3.2.1 Macroeconomic factors and data

Comparing with the default rate of Chinese joint-equity commercial bank, we use GDP, GDP Growth, CPI, financial expenditure (FE), unemployment rate (UR), interest rate (IR), value of exports (VE), value of shares (VS), exchange rate(dollar) (ER), and disposable income (DI) to be the macroeconomic factors.
WE also use time series data on a quarter base over the years 2009-2013 and we also choose the probability of the non-performing loan to be the default rate. The data of the non-performing loan is from the official website of De centrale bank van Nederland[18], and the data of all the macroeconomic factors is from the official website of Centraal Bureau voor de Statistiek [17].
Also by the CPV model, $P_t = \frac{1}{1+e^{-Y_t}}$,we can get Y_t for every quarter. The results are in the table below.
In the table we have made no seasonal adjustment and CPI index as the provided has already been adjusted for seasonal influences.

Table 3.10: *All of the required data*

DR	Y	GDP	CPI	Growth	UR	FE	IR	VE	ER	VS	DI
0.0183	-3.98	136125	107.38	-0.020838	2.2	71107	3.74	45413.75	1.61	255493	57327
0.0244	-3.69	134183	107.39	-0.01427	2.5	74446	3.86	45853.63	1.65	291680	79723
0.0269	-3.59	135242	106.46	0.007892	2.6	71926	3.65	50718.59	1.67	351963	58309
0.0320	-3.41	135794	105.82	0.004082	2.9	77303	3.5	53809.24	1.7	383486	63513
0.0319	-3.41	136537	108.08	0.005472	3.3	73092	3.4	54976.71	1.68	400607	57362
0.0276	-3.56	136999	107.64	0.003384	3.1	79490	3.08	52906.83	1.71	382359	79565
0.0257	-3.64	137197	107.96	0.001445	2.7	71289	2.65	55473.88	1.76	399374	61742
0.0282	-3.54	138552	107.77	0.009876	2.6	77413	2.84	60210.07	1.67	423867	63611
0.0273	-3.57	139360	110.08	0.005832	2.9	72761	3.35	63797.67	1.68	438484	59904
0.0268	-3.59	139148	110.12	-0.00152	2.6	78241	3.44	66546.33	1.6	408607	82553
0.0272	-3.58	138698	111.15	-0.00323	2.6	71338	2.73	66371.52	1.53	356411	60555
0.0271	-3.58	137696	110.48	-0.00722	2.8	76375	2.43	62060.34	1.57	393273	64634
0.0294	-3.50	137315	113.26	-0.00277	3.2	73558	2.23	64217.45	1.67	411636	60075
0.0312	-3.43	137929	112.87	0.004471	3.1	79117	2.06	63202.37	1.68	400283	81077
0.0306	-3.45	136731	113.98	-0.00869	3.1	71953	1.78	61126.92	1.8	427506	61504
0.0310	-3.44	135919	114.2	-0.00594	3.3	77461	1.66	63279.58	1.62	438103	64812
0.0278	-3.55	135414	116.88	-0.00372	4	70224	1.74	65103.45	1.51	456433	59752
0.0300	-3.48	135191	116.44	-0.00165	4.1	79366	1.78	63306.37	1.51	459106	80232
0.0295	-3.49	135929	116.7	0.005459	4.1	72934	1.66	64655.91	1.57	492268	62661
0.0323	-3.40	136887	115.81	0.007048	4.7	77505	1.74	64989.57	1.69	507518	67544

3.2.2 Model building

In CPV model,Y_t is an index value derived using a multi-factor regression model that considers a number of macro economic factors, where t is the time period.

$$Y_t = \beta_0 + \sum_{k=1}^{K} \beta_k X_{k,t} + \varepsilon_{k,t},$$

So in this case study

$$Y_t = \beta_0 + \beta_1 CPI + \beta_2 GDP + \beta_3 Growth + \beta_4 UR + \beta_5 FE + \beta_6 DI + \beta_7 VE + \beta_8 ER + \beta_9 VS + \beta_0 IR$$

Table 3.11: *The regression results*

	Coefficients	Std. Error	t value	$Pr(>\mid t \mid)$
(Intercept)	6.60E+00	3.72E+00	1.774	0.10982
GDP	-8.71E-05	2.15E-05	-4.059	0.00285
CPI	-2.40E-02	2.09E-02	-1.149	0.2802
Growth	4.74E+00	3.52E+00	1.346	0.21121
UR	9.17E-02	7.43E-02	1.233	0.24871
FE	1.38E-05	9.19E-06	1.505	0.1667
DI	-2.36E-06	3.02E-06	-0.783	0.4538
IR	6.41E-03	5.43E-02	0.118	0.9085
VE	3.65E-05	8.81E-06	4.145	0.0025
ER	9.91E-01	2.70E-01	3.675	0.00511
VS	-1.33E-06	8.84E-07	-1.503	0.16712
Multiple R-squared:	0.9061	Adjusted R-squared	0.8018	
F-statistic:	8.689	p-value	0.001643	

In this regression results table 3.11, the R-squared is 0.9061, Adjusted R-squared is 0.8018, F-statistic is 8.689. P-value is 0.001643. This means that the hypothesis H_0 :" all regression coefficients are zero" is strongly rejected. That is there is explanatory power in this model. But in several individual t-tests the p-value are large. As mentioned in Section 3.12 this could be due to multi-collinearity or due to lack of influence on Y_t.
The method I will use next is the backward elimination procedure.

Backward elimination procedure

Table 3.12: *Backward elimination procedure table*

Start AIC=-107.61	Df	Sum of Sq	RSS	AIC
IR	1	0.000048	0.030711	-109.578
DI	1	0.002088	0.032751	-108.291
none	1		0.030663	-107.609
CPI	1	0.004497	0.03516	-106.871
UR	1	0.005182	0.035845	-106.486
Growth	1	0.006173	0.036836	-105.94
VS	1	0.007695	0.038358	-105.13
FE	1	0.007712	0.038375	-105.122
ER	1	0.046019	0.076682	-91.276
GDP	1	0.056131	0.086794	-88.799
VE	1	0.058523	0.089186	-88.255
Step:AIC=-109.58				
	Df	Sum of Sq	RSS	AIC
DI	1	0.002328	0.033038	-110.116
none			0.01095	-130.2
Growth	1	0.006138	0.036848	-107.933
UR	1	0.006204	0.036914	-107.898
VS	1	0.007801	0.038512	-107.05
CPI	1	0.0095	0.04021	-106.187
FE	1	0.009567	0.040278	-106.154
ER	1	0.05324	0.083951	-91.465
GDP	1	0.057926	0.088637	-90.379
VE	1	0.058523	0.089234	-90.245
Step:AIC=-110.12				
	Df	Sum of Sq	RSS	AIC
none			0.033038	-110.116
Growth	1	0.005289	0.038328	-109.146
VS	1	0.006363	0.039401	-108.594
UR	1	0.007125	0.040163	-108.211
FE	1	0.010228	0.043267	-106.722
CPI	1	0.013641	0.04668	-105.204
ER	1	0.052847	0.085885	-93.01
GDP	1	0.058459	0.091497	-91.744
VE	1	0.060749	0.093788	-91.249

The AIC is used for backward elimination. AIC = 2 log(likelihood) + 2p with p the number of parameters in the model, smaller values point to better fitting models. Each variable is removed from the model in turn, and the resulting AIC's are reported. For eight regressors the AIC deteriorates (becomes larger) by removal, so these variables are important. For two regressors removal makes the AIC smaller (better), so these regressors are candidates for removal. After we remove them the model is $Y_t = \beta_0 + \beta_1 CPI + \beta_2 GDP + \beta_3 Growth + \beta_4 UR + \beta_5 FE + \beta_6 VE + \beta_7 ER + \beta_8 VS$ The regression results is in the table below

In the regression results table 3.13, the R-squared is 0.8989, Adjusted R-squared is0.8253, F-statistic is 12.22. P-value is 0.0001778. It also means that H_0 :" all regression coefficients are zero" is strongly rejected, hence there is explanatory power in this model. But for the individual t-tests, the p-values of GDP Growth, and value of shares, are still big.
We try to remove them and build a new model,

$$Y_t = \beta_0 + \beta_1 CPI + \beta_2 GDP + \beta_3 UR + \beta_4 VE + \beta_5 ER$$

The regression results is in the table 3.14.

Table 3.13: *regression results after removing IR and DI*

	Coefficients	Std. Error	t value	$Pr(>\mid t \mid)$
(Intercept)	7.38E+00	3.13E+00	2.361	0.037751
data1$GDP	-8.69E-05	1.97E-05	-4.412	0.001043
data1$CPI	-2.94E-02	1.38E-02	-2.131	0.056467
data1$Growth	4.31E+00	3.25E+00	1.327	0.211386
data1$UR	1.01E-01	6.54E-02	1.54	0.151774
data1$FE	7.93E-06	4.30E-06	1.845	0.09205
data1$VE	3.71E-05	8.25E-06	4.497	0.000905
data1$ER	9.75E-01	2.32E-01	4.195	0.001499
data1$VS	-1.18E-06	8.09E-07	-1.455	0.173475
Multiple R-squared:	0.8989	Adjusted R-squared	0.8253	
F-statistic:	12.22	p-value	0.0001778	

Table 3.14: *regression results after removing IR,DI,Growth and VS*

	Coefficients	Std. Error	t value	$Pr(>\mid t \mid))$
(Intercept)	8.476e+00	3.191e+00	2.656	0.01879
data1$GDP	-8.185e-05	2.104e-05	-3.890	0.00163
data1$CPI	-4.121e-02	9.892e-03	-4.167	0.00095
data1$UR	1.014e-01	4.026e-02	2.518	0.02461
data1$VE	3.489e-05	6.232e-06	5.599	6.56e-05
data1$ER	8.318e-01	2.091e-01	3.979	0.00137
Multiple R-squared:	0.8441	Adjusted R-squared	0.7884	
F-statistic:	15.16	p-value	3.201e-05	

In the table we can also see that the p-values of all the factors are not too large. So the final model is

$$Y_t = \beta_0 + \beta_1 CPI + \beta_2 GDP + \beta_3 UR + \beta_4 FE + \beta_5 VE + \beta_6 ER$$

According to the table of The regression results of final model, we can get the model of Y_t,

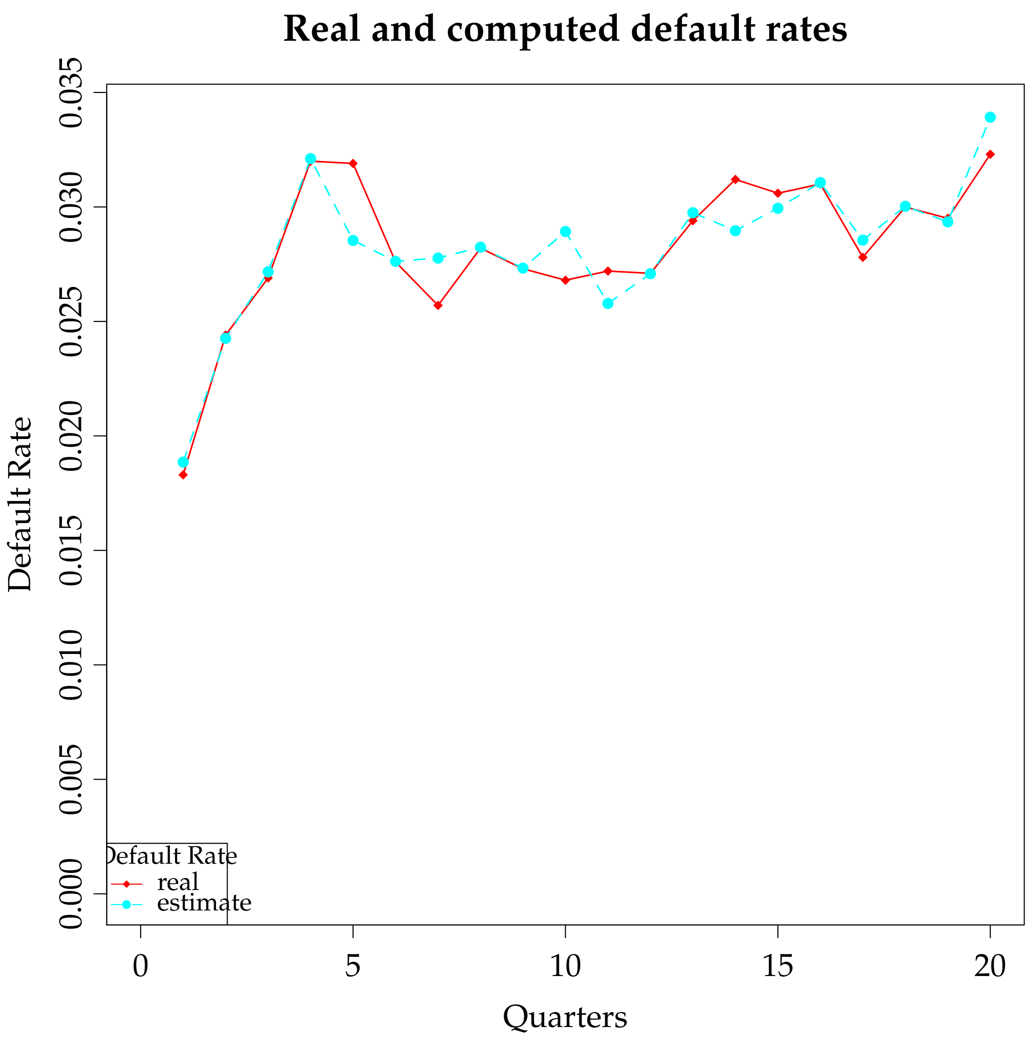
$$Y_t = 8.476 - 0.00008185GDP - 0.04121CPI + 0.1014UR + 0.00003489VE + 0.8318ER$$

3.2.3 Calculating the default rate

With the formula for Y_t obtained in Section 3.2.2, we compute the default rates and compare them with the real default rates in table 3.15 and the picture below.

Table 3.15: *comparing with real default rate and estimate default rate*

real default rate	estimate default rate
0.0183	0.0190
0.0244	0.0240
0.0269	0.0277
0.0320	0.0319
0.0319	0.0292
0.0276	0.0268
0.0257	0.0285
0.0282	0.0279
0.0273	0.0279
0.0268	0.0283
0.0272	0.0265
0.0271	0.0268
0.0294	0.0300
0.0312	0.0280
0.0306	0.0302
0.0310	0.0303
0.0278	0.0296
0.0300	0.0291
0.0295	0.0298
0.0323	0.0339



3.2.4 Conclusion and discussion

The plot above shows that the model gives a fairly good fit to the data. As in Section 3.1.4 we also evaluate the prediction quality of the model by predicting the default of the 20th quarter by means of the model with the coefficients estimated based on the first 19 quarters.
After we get the final model, we want to know how good the model is. We use the first 19 quarters of historic data to fit the parameters and predict the 20th quarter's default rate. Then we compare this result with the real default rate of the 20th quarter. Also we compare the result with "tomorrow is same as today" prediction, and find which method is better.

Table 3.16: *regression results according to the first 19 quarters*

	Coefficients	Std. Error	t value	$Pr(>\mid t\mid))$
(Intercept)	7.118e+00	3.338e+00	2.132	0.052619
data1$GDP	-7.201e-05	2.227e-05	-3.233	0.006533
data1$CPI	-4.145e-02	9.739e-03	-4.256	0.000936
data1$UR	1.411e-01	5.161e-02	2.735	0.017024
data1$VE	3.275e-05	6.388e-06	5.127	0.000194
data1$ER	8.598e-01	2.071e-01	4.152	0.001138
Multiple R-squared:	0.8497	Adjusted R-squared	0.7918	
F-statistic:	14.7	p-value	5.885e-05	

According to the table 3.16 of The regression results, based on the first 19 quarters, we can get the model of Y_t.

$$Y_t = 7.118 - 0.000072GDP - 0.04145CPI + 0.1411UR + 0.00003275VE + 0.8598ER$$

We put the macroeconomic historic data of the 20th quarter into the model above and we can easily get the estimated default rate of the 20th quarter is 0.0357. Comparing this estimated default rate with the real default 0.0323 we can see the difference is very close. Moreover, if we compare the estimated default rate 0.0357 with the real default rate of the 19th quarter 0.0295, we can find the estimated default rate of the 20th quarter is much closer to the real default rate of the 20th quarter. In this case the prediction by the model is much better than predicting by the value of the previous quarter.

The prediction made by the model turns out to be very good. To be convinced of the quality of the model we would need more good predictions. As discussed in Section 3.1.4, the set of the 20 data points, however, is too small for a good estimate of the parameters and enough data points left to compare the predictions.

Comparing CPV model for Chinese data and Dutch data

If we compare the model for the Chinese data (Section 3.1) and the Dutch data (Section 3.2), we observe that in both cases the model fits well to the data, although the fit for the Dutch data is not as good as for the Chinese data. Concerning the prediction quality of the model it seems to be the other way around. As pointed out earlier, a full evaluation of the prediction quality needs more data.

The models for the Chinese and Dutch data are quite different with respect to the macro-economic factors that appear in the formula for Y_t. Since the nature of the Chinese and Dutch economies are very different, it is not surprising that different macro-economic factors influence the default rate.

Chapter 4

Estimation Default Risk of Both Chinese and Dutch Companies Based on KMV Model

4.1 Use KMV model to evaluate default risk for CNPC and Sinopec Group

We will use the KMV model, which is explained in Section 2.2.2, to evaluate the default risk of two Chinese oil companies: CNPC and Sinopec Group. We preform the steps described in Section 2.2.2, to compute the default distances. Moreover, we will compare the default distances of CNPC and Sinopec Group.

1.Data Source
Our study sample are the financial data of the largest two petrochemical company of China (CNPC and Sinopec Group) from the second quarter of 2012 to the third quarter of 2013. The related data are: interest rate, daily stock closing price, the market value, short-term liabilities and long-term liabilities. In the daily stock closing prices we only take the available prices of the days that the stock market is open. The data is from the website of the Netease Finance.[19]

2.The Market value
The market value of the two companies are shown in the table 4.1

3.Default Point Calculation

Table 4.1: *The market value of CNPC and Sinopec Group*

	CNPC	Sinopec Group
2012Q3	1.66915E+12	5.42626E+11
2012Q4	1.60692E+12	5.20053E+11
2013Q1	1.65451E+12	6.04269E+11
2013Q2	1.39279E+12	4.97734E+11
2013Q3	1.43488E+12	5.1755E+11

According to the KMV model the default point satisfies $DP = STD + 0.5LTD$, where STD is short-term liabilities, and LTD is long-term liabilities. The DP of the two companies are shown in the table 4.2.

Table 4.2: *The default point of CNPC and Sinopec Group(Million Yuan)*

	CNPC	Sinopec Group
2012Q3	772965	546573
2012Q4	781410	594890
2013Q1	835995	614902
2013Q2	863635	598605
2013Q3	903328	586453

4.Asset value and Asset Value Fluctuation Ration Calculation
We use historical stock closing price data to calculate the stock fluctuation ratio σ_s, assuming the historical data fit the log-normal distribution. The daily logarithmic profit ratio is

$$u_i = ln(\frac{S_i}{S_{i-1}}),$$

where S_i is the daily stock closing price of day i. So the stock fluctuation ratio in daily stock returns is:

$$S = \sqrt{\frac{1}{n-1}\sum_{1}^{n-1}(u_i - \bar{u})},$$

where $\bar{u}$ is the mean of u_i. The number of trading days quarterly of the stock is N, so the relationship between the quarterly fluctuation ratio σ_s and daily fluctuation ratio S is

$$\sigma_s = S\sqrt{N}.$$

The stock fluctuation ratio σ_s of the two companies are shown in the table below.

Table 4.3: *The stock fluctuation ratio σ_s of CNPC and Sinopec Group*

	CNPC	Sinopec Group
2012Q3	0.070034545	0.115095849
2012Q4	0.069165841	0.097894454
2013Q1	0.069165841	0.115935752
2013Q2	0.068029746	0.317494085
2013Q3	0.070759383	0.115512544

According to the formula above we can estimate the asset value and its volatility. We use matlab 2012b to solve the nonlinear equations. The code is straightforward and may be found on the world wide web.

```
function F=KMVfun(EtoD,r,T,EquityTheta,x)
d1=(log(x(1)*EtoD)+(r+0.5*x(2)^2*T))/x(2);
d2=d1-x(2);
F=[x(1)*normcdf(d1)-exp(-r)*normcdf(d2)/EtoD-1;normcdf(d1)*x(1)*x(2)
-EquityTheta];
end
function [Va,AssetTheta]=KMVOptSearch(E,D,r,T,EquityTheta)
EtoD=E/D;
x0=[1,1];
VaThetax=fsolve(@(x)KMVfun(EtoD,r,T,EquityTheta,x),x0);
Va=VaThetax(1)*E;
AssetTheta=VaThetax(2);
end
```

The relative asset value and its volatility are shown in the table below.

Table 4.4: *Asset value of CNPC and Sinopec Group(Million Yuan)*

	CNPC	Sinopec Group
2012Q3	2420000	1070000
2012Q4	2370000	1100000
2013Q1	2470000	1200000
2013Q2	2230000	1080000
2013Q3	2310000	1090000

Table 4.5: *The asset volatility of CNPC and Sinopec Group*

	CNPC	Sinopec Group
2012Q3	0.0483	0.0582
2012Q4	0.047	0.0464
2013Q1	0.0464	0.0583
2013Q2	0.0425	0.1465
2013Q3	0.0439	0.055

5.Find the default distance(DD)

At last, according to the formula above, we can find DD of the three firms, we also use matlab 2012b to do the calculation. The matlab code are:

```
function F=DDfun(Va,AssetTheta,D)
F=[(Va−D)/(Va∗AssetTheta)];
```

The relative results are shown in the table below.

Table 4.6: *The default distance of CNPC and Sinopec Group*

	CNPC	Sinopec Group
2012Q3	14.0832	8.4298
2012Q4	14.4301	9.8697
2013Q1	14.2421	8.3661
2013Q2	14.4301	3.0377
2013Q3	13.8695	8.3672

We calculate the asset value, asset volatility and default distance in the second quarter of 2012 of CNPC as an example, The Matlab code is:

```
>> r=0.03;
T=1;
E=1.66915E+12;
D=7.72965E+11;
EquityTheta=0.070034545;
[Va,AssetTheta]=KMVOptSearch(E,D,r,T,EquityTheta)
[DD]=DDfun(Va,AssetTheta,D)

Equation solved.

fsolve completed because the vector of function values is near zero
```

```
as measured by the default value of the function tolerance, and
the problem appears regular as measured by the gradient.

<stopping criteria details>


Va =

   2.4193e+12


AssetTheta =

    0.0483


DD =

   14.0832
```

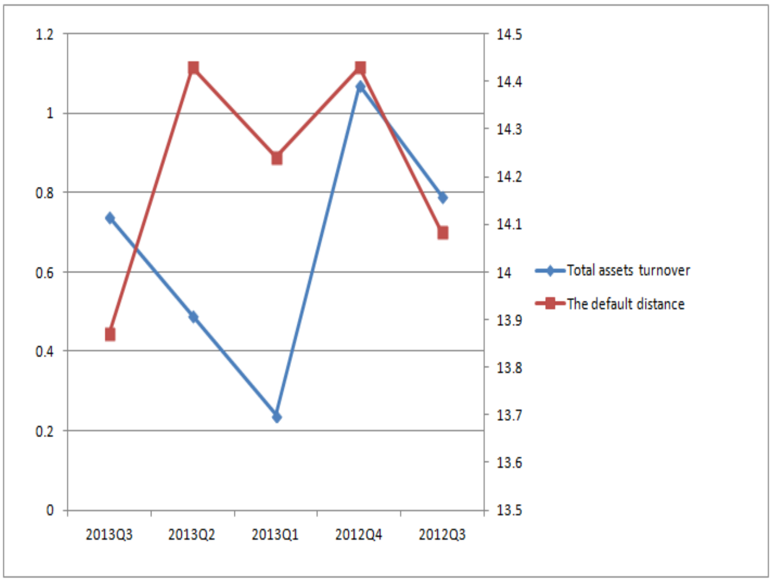
The results of the code above shows that the asset value, asset volatility and default distance of CNPC in the second quarter of 2012. We can change the Initialize variables and repeat the procedure to calculate the asset value, asset volatility and default distance for other time period and companies.

6.Comparing the default distance with the total assets turnover
Sometimes total asset turnover is considered to be a measure for the default risk of a company. We compare the total assets turnover with the default distance, both for CNPC and Sinopec Group. The Total assets turnover of the two companies are shown in the table below.

Table 4.7: *The Total assets turnover of CNPC and Sinopec Group*

	CNPC	Sinopec Group
2012Q3	0.79	1.75
2012Q4	1.07	2.34
2013Q1	0.24	0.55
2013Q2	0.49	1.12
2013Q3	0.74	1.67

Comparison of Total assets turnover and default distance of CNPC



Comparison of Total assets turnover and default distance of Sinopec Group.

Except for the third quarter of 2013 for CNPC, the trend in the total assets turnover and the default distance are the same in both cases in all other quarters, which is as expected.

4.2 Use KMV model to evaluate default risk for Royal Dutch Shell and Royal Philips

We perform the same analysis as in Section 4.1, but now for two large Duthc companies: Royal Dutch Shell and Royal Philips.

1.Data source
Our study sample are the financial data of two Dutch companies,Royal Dutch Shell and Royal Philips from the second quarter of 2012 to the third quarter of 2013. The related data are: interest rate, daily stock closing price, the market value, short-term liabilities and long-term liabilities. The data is from the official webset of the two companies and YAHOO Fiance. [20]

2.The Market value
The market value of the two companies are shown in the table below

Table 4.8: *Market Value of Royal Dutch Shell and Royal Philips*

	Royal Dutch Shell	Royal Philips
2012Q3	2.26E+11	2.18E+10
2012Q4	2.24E+11	2.43E+10
2013Q1	2.11E+11	2.70E+10
2013Q2	2.10E+11	2.46E+10
2013Q3	2.19E+11	2.95E+10

3.Default Point Calculation
According to the KMV model we have $DP = STD + 0.5LTD$, where DP is the default point, STD is short-term liabilities and LTD is long-term liabilities. The DP of the two companies are shown in the table below.

Table 4.9: *Default Point of Royal Dutch Shell and Royal Philips*

	Royal Dutch Shell	Royal Philips
2012Q3	1.39698E+11	13384500000
2012Q4	1.33689E+11	13925500000
2013Q1	1.37688E+11	13081000000
2013Q2	1.29279E+11	13119500000
2013Q3	1.33995E+11	12752500000

4.Asset value and Asset Value Fluctuation Ration Calculation
We use historical stock closing price data to calculate the stock fluctuation ratio σ_s, assuming the historical data fit the log-normal distribution. The daily logarithmic profit ratio is

$$u_i = ln(\frac{S_i}{S_{i-1}}),$$

where S_i is the relative daily stock closing price.
As before, fluctuation ratio in daily stock returns is:

$$S = \sqrt{\frac{1}{n-1}\sum_1^{n-1}(u_i - \bar{u})},$$

where $\bar{u}$ is the mean of u_i. The relationship between the quarterly fluctuation ratio σ_s and daily fluctuation ratio S is

$$\sigma_s = S\sqrt{N},$$

The stock fluctuation ratio σ_s of the two companies are shown in the table below.

Table 4.10: *The volatility of equity of Royal Dutch Shell and Royal Philips*

	Royal Dutch Shell	Royal Philips
2012Q3	0.062748481	0.09851447
2012Q4	0.073478442	0.102124553
2013Q1	0.056810285	0.12547093
2013Q2	0.069778976	0.088840043
2013Q3	0.077046101	0.09262708

According to the formula above we can find the asset value and its volatility. We use matlab 2012b to solve the nonlinear equations, The relative asset value and its volatility are shown in the table 4.11.

5.Find the default distance(DD)
At last, according to the formula above, we can find DD of the two firms, we also use matlab 2012b to do the calculation. The relative results are shown in the table below.

If we compare the default distance of the two Chinese and the two Dutch companies as given in tables 4.6 and 4.13, we see that, on average,

Table 4.11: *Asset value of Royal Dutch Shell and Royal Philips*

	Royal Dutch Shell	Royal Philips
2012Q3	3.63E+11	3.49E+10
2012Q4	3.56E+11	3.75E+10
2013Q1	3.46E+11	3.99E+10
2013Q2	3.37E+11	3.75E+10
2013Q3	3.45E+11	4.20E+10

Table 4.12: *Asset volatility of Royal Dutch Shell and Royal Philips*

	Royal Dutch Shell	Royal Philips
2012Q3	0.0391	0.0615
2012Q4	0.0463	0.0662
2013Q1	0.0346	0.085
2013Q2	0.0435	0.0583
2013Q3	0.0489	0.0651

Table 4.13: *Default Distance of Royal Dutch Shell and Royal Philips*

	Royal Dutch Shell	Royal Philips
2012Q3	15.757	10.0372
2012Q4	13.4837	9.7085
2013Q1	17.4122	7.906
2013Q2	14.1563	11.1373
2013Q3	12.8035	10.689

Royal Dutch Shell has the highest default distance, so the least default risk. Over time, the precise value of the default distance fluctuates, where the value for CNPC seems most stable. The default distance seems to give a good general impression about the default risk of a company.

Chapter 5

Prediction of default risk of a portfolio

The previous chapters discuss various ways to model and estimate the default risk of a single company. Most investors, in particular banks and insurance companies, lend money to many companies and individuals. It is important for them to understand the risk profile of their entire portfolio of loans. This risk profile is determined by the risks of the individual obligors and their correlations. In this chapter we study the influence of correlations between obligors on the default risk of a portfolio.

5.1 Two models

We will consider two models for default risk in a portfolio of loans. In the first model there are no correlation and in the second model there are correlations. We consider N obligors, the first model will be a Bernoulli model, where to each obligor $i \in \{1,...,N\}$ we assign a random variable L_i which is 1 in case of default and zero otherwise. All these random variable are considered to be identically distributed and independent.
The second model will be a factor model. We assign to each obligor $i \in \{1,...,N\}$ a standard normal random variable r_i, which is the sum of a normal random variable $R\Phi$, where Φ is standard normal and $R \in [0,1]$, that is the same for all obligors and an independent normal random variable ϵ_i. The ϵ_i are assumed to be independent. Default of obligor i is assumed to occur when r_i will be lower than a certain constant $-c$ with $c \geq 0$. The default will be correlated due to the common random variable Φ.
In fact for both models we will consider a range of time periods, say

months or quarters. In each such a time period we will consider defaults of the obligors as described by the two models and we will assume that all random variables belonging to different time periods are independent.
Consider a bank with a portfolio of loans. It is important for the bank to predict the expected losses due to the defaults in the next time period. These prediction will be based on historic data, by using a suitable model. We will study how the models without correlations and the model with correlations can be used for such predictions. Moreover we will investigate the precision of such prediction. In particular we want to study the effect of ignoring correlation to the quality of the prediction. We will do so by considering two data sets: one generated by the Bernoulli model and one generated by the factor model. Based on the both datasets we will estimate the default probability and the probability that more than p percent of the portfolio will be in default for various p. These estimate will be compared to the actual rates in the datasets.
We will use both models for both datasets. In addition, we will consider a third dataset generated by a two factor model and evaluate the quality of predictions made by the one factor model.

5.1.1 The Uniform Bernoulli Model

Let us assume that
1.All loans are the same amount, which we scale to be 1.
2.All obligors have same default probability p per year.
3.All obligors are independent.
Then the model becomes a Uniform Bernoulli Model,

$$L_i \sim B(1;p), \ \ i.e., L_i = \begin{cases} 1 \ \text{ with probability } p, \\ 0 \ \text{ with probability } 1-p, \end{cases}$$

$$L_i, \qquad i = 1, ..., N \ \text{ independent.}$$

Note that the expected number of defaults is $p * N$
According to the Maximum Likelihood Estimator we can estimate p in year i by means of the historic data,

$$\hat{p}_i = \frac{\text{number of defaults in year i}}{N} i = 1, ..., T$$

,and average over all years by

$$\hat{p} = 1/T(\hat{p}_1 + ... + \hat{p}_T)$$

,where T is the number of years in historic data.

5.1.2 Factor Model

The assumption that all obligors are independent is not very realistic. It may be expected that the financial strength of obligor depends on certain factors in the economy. If a group of obligors depends of the same factor, they will be correlated through that factor. The factor model captures this feature.

Assume that the set of obligors is divided into two groups:
$1,\ldots,N = I_1 \cup I_2 \cup \ldots \cup I_k$, where $I_i \cup I_k = \varnothing$. We view the set I_i as all obligors of factor type i. We also assume all obligors of type i have the default probability p_i, and all obligors are independent.

Then we can use Factor model to estimate the default probability (see also Section 1.2.3 of [2]).

Based on Merton's model, in each year obligor i has a log asset value:

$$r_i(t) = R_i\Phi(t) + \varepsilon_i(t)$$

, where$R_i \in [-1,1]$,$\Phi(t) \sim N(0,1)$,$\varepsilon_i(t) \sim N(0,1-(R_i)^2)$, $\Phi(t)$ and $\varepsilon_i(t)$ independent.

Then

$$R_i\Phi(t) \sim N(0,R_i^2),$$

so

$$\mathbb{E}r_i(t) = R_i\mathbb{E}\Phi(t) + \mathbb{E}\varepsilon_i(t) = 0,$$

and,

$$\begin{aligned}
\mathbb{V}r_i(t) &= \mathbb{E}(r_i(t))^2 \\
&= \mathbb{E}(R_i\Phi(t) + \varepsilon_i(t))^2 \\
&= \mathbb{E}(R_i\Phi(t))^2 + 2\mathbb{E}R_i\Phi(t)\varepsilon_i(t) + \mathbb{E}(\varepsilon_i(t))^2 \\
&= R_i^2 + (1 - R_i^2) = 1.
\end{aligned}$$

So

$$r_i(t) \sim N(0,1)$$

Obligor i is default if $r_i(t) < -c$, this happens with probability

$$\mathbb{P}(r_i(t) < -c) = F(-c) = p$$

where

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-\frac{y^2}{2})dy.$$

If all $R_i = 0$, then all of the obligors are independent.

If all $R_i = 1$, then all of the obligors are completely dependent: all are in default or not.

Computing more than 10 percent of the portfolio defaults next year

The probability that exactly k of N obligors in default is

$$\begin{aligned} &\mathbb{P} \quad (\text{exactly k of N obligors in default}) \\ &= \quad \mathbb{E}\mathbb{P}(\text{exactly k of N obligors in default} \mid \Phi(t) = S). \end{aligned}$$

Given that

$$\Phi(t) = S,$$

Obligor 1 is in default precisely when

$$r_1(t) < -c.$$

So

$$RS + \varepsilon_1(t) < -c,$$

which is equivalent to

$$\frac{\varepsilon_1(t)}{\sqrt{1-R^2}} < \frac{-c-RS}{\sqrt{1-R^2}}.$$

Hence,

$$\mathbb{P}(\text{obligor 1 in default} \mid \Phi = S) = F(\frac{-c-RS}{\sqrt{1-R^2}}).$$

Since,$\frac{\varepsilon_1(t)}{\sqrt{1-R^2}}$ for $i = 1,...N$ are independent $N(0,1)$ variable we find

$$\begin{aligned} &\mathbb{P} \quad (\text{exactly k of N obligors in default} \mid \Phi(t) = S) \\ &= \quad \binom{N}{k} F(\frac{-c-RS}{\sqrt{1-R^2}})^k (1 - F(\frac{-c-RS}{\sqrt{1-R^2}}))^{N-k}. \end{aligned}$$

Thus, taking expectation over the values S,

$$\begin{aligned} &\mathbb{P} \quad (\text{exactly k of N obligors in default}) \\ &= \quad \binom{N}{k} \int_R F(\frac{-c-RS}{\sqrt{1-R^2}})^k (1 - F(\frac{-c-RS}{\sqrt{1-R^2}}))^{N-k} \frac{1}{\sqrt{2\pi}} exp(-\frac{S^2}{2}) dS. \end{aligned}$$

So,

$$\begin{aligned} &\mathbb{P} \quad (\text{more than 10 percent of obligors in default}) \\ &= \quad 1 - \mathbb{P}(\text{less than } \frac{N}{10} - 1 \quad \text{obligors in default}) \\ &= \quad 1 - \sum_{k=0}^{\frac{N}{10}-1} \binom{N}{k} \int_R F(\frac{-c-RS}{\sqrt{1-R^2}})^k (1 - F(\frac{-c-RS}{\sqrt{1-R^2}}))^{N-k} \frac{1}{\sqrt{2\pi}} exp(-\frac{S^2}{2}) dS. \end{aligned}$$

Estimate the R_i and $-c$ from the historic data

Now we need to estimate the R_i and $-c$ from the historic data. The bank does not observe r_i directly, but only $r_i < -c$ or not.
For estimating $-c$, we have

$$\mathbb{P}(r_i(t) < -c) = p,$$

and we can estimate p by using

$$\hat{p} = \frac{\text{Number of default obligors}}{\text{Total number of obligors}}.$$

So $-c$ is the "quantile" of the normal distribution corresponding to probability p.
Next we consider estimating the coefficients R_i. We will use the estimator $\hat{V}$ for $\mathbb{E}[\text{number of defaults}^2]$ given by $\hat{V} = (\text{Number of defaults})^2$. We have

$$\begin{aligned} &\mathbb{P} \quad (\text{obligor 1 and obligor 2 both default}) \\ = \;& \mathbb{P}(r_1(t) < -c \quad \text{and} \quad r_2(t) < -c) \\ = \;& \mathbb{E}[\mathbb{P}(R_1\Phi(t) + \varepsilon_1(t) < -c \quad \text{and} \quad R_2\Phi(t) + \varepsilon_2(t) < -c|\Phi(t) = S)]. \end{aligned}$$

Also,

$$\begin{aligned} &\mathbb{P}(\varepsilon_1(t) < -c - R_1 S \quad \text{and} \quad \varepsilon_2(t) < -c - R_2 S) \\ = \;& \mathbb{P}(\varepsilon_1(t) < -c - R_1 S)\mathbb{P}(\varepsilon_2(t) < -c - R_2 S) \\ = \;& \mathbb{P}(\frac{\varepsilon_1(t)}{\sqrt{1-R_1^2}} < \frac{-c - R_1 S}{\sqrt{1-R_1^2}})\mathbb{P}(\frac{\varepsilon_2(t)}{\sqrt{1-R_2^2}} < \frac{-c - R_2 S}{\sqrt{1-R_2^2}}) \\ = \;& F(\frac{-c - R_1 S}{\sqrt{1-R_1^2}})F(\frac{-c - R_2 S}{\sqrt{1-R_2^2}}). \end{aligned}$$

Hence,

$$\begin{aligned} &\mathbb{P}(r_1(t) < -c \quad \text{and} \quad r_2(t) < -c) \\ = \;& \mathbb{E}\mathbb{P}[R_1\Phi(t) + \varepsilon_1(t) < -c \quad \text{and} \quad R_2\Phi(t) + \varepsilon_2(t) < -c \mid \Phi = S] \\ = \;& \mathbb{E}[F(\frac{-c - R_1\Phi}{\sqrt{1-R_1^2}})F(\frac{-c - R_2\Phi}{\sqrt{1-R_2^2}})]. \end{aligned}$$

Let us make one more assumption to simplify the model. We will assume that all obligors have the same coefficient of dependence of $\Phi(t)$. In other

words, all the R_i are the same, so $R_i = R$ for all i.
Then,

$$\begin{aligned} & \mathbb{E}F(\frac{-c - R_1\Phi(t)}{\sqrt{1-R_1^2}})F(\frac{-c - R_2\Phi(t)}{\sqrt{1-R_2^2}}) \\ = & \mathbb{E}F(\frac{-c - R\Phi(t)}{\sqrt{1-R^2}})F(\frac{-c - R\Phi(t)}{\sqrt{1-R^2}}) \\ = & \int_{\mathbb{R}} F(\frac{-c - RS}{\sqrt{1-R^2}})F(\frac{-c - RS}{\sqrt{1-R^2}})\frac{1}{\sqrt{2\pi}}\exp(-\frac{S^2}{2})dS. \end{aligned}$$

Consider one year situation:

$$L_i = \begin{cases} 1 & r_i < -c, \text{obligor i in default} \\ 0 & r_i \geq -c, \text{obligor i not in default} \end{cases}.$$

We have

$$\mathbb{E}[(L_1 + \cdots + L_N)^2] = \mathbb{E}\sum_{i=1}^{N}\sum_{j=1}^{N} L_i L_j = \sum_{i=1}^{N}\sum_{j=1}^{N} \mathbb{E}L_i L_j.$$

For $i \neq j$ we have,

$$\mathbb{E}L_i L_j = \mathbb{E}1_{r_i<-c}1_{r_j<-c} = \mathbb{E}1_{r_i<-c \text{ and} r_j<-c}$$

and

$$\begin{aligned} & \mathbb{P} \quad (r_i < -c \text{ and} r_j < -c) \\ = & \int_R F(\frac{-c - R_1 S}{\sqrt{1-R_1^2}})^2 \frac{1}{\sqrt{2\pi}}\exp^(-\frac{S^2}{2})ds. \end{aligned}$$

For $i = j$ we have,

$$\mathbb{E}L_i L_j = \mathbb{E}L_i^2 = p.$$

Hence

$$\mathbb{E}[(L_1 + \cdots + L_N)^2] = N(N-1)H_c(R) + Np,$$

where,

$$H_c(R) = \int_R F(\frac{-c - R_1 S}{\sqrt{1-R_1^2}})^2 \frac{1}{\sqrt{2\pi}}\exp(-\frac{S^2}{2})ds.$$

Since $\mathbb{E}[(L_1 + \cdots + L_N)^2]$ is estimated by$\hat{V}$, and c by $\hat{c}$, we obtain the following estimate for R:

$$\hat{R} = H_{\hat{c}}^{-1}(\frac{\hat{V} - N\hat{p}}{N(N-1)})$$

5.2 Computer simulation

We will now generate a dataset without correlations and a dataset with correlations and use the Bernoulli model and the factor model to estimate default risk. As measures of default risk we consider the default probability and the probability that more than q percent of the portfolio will be in default, for some suitable value of q.

5.2.1 Simulation of the Uniform Bernoulli Model

The simulation procedure of the Uniform Bernoulli Model are:

1. Data Production

We take the number of obligors $N = 1000$, time duration $T = 20$(years), default probability for each obligor is $p = 0.05$. Because this is Uniform Bernoulli Model, we assume each obligor has the same default probability. So for each year, we generate N independent Bernoulli variables with parameter p.

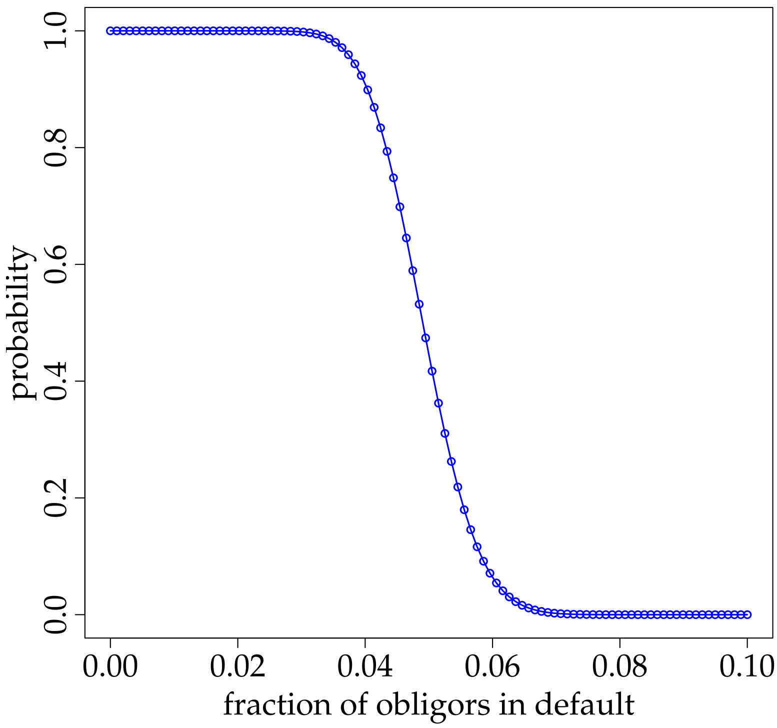
2. Estimate the next time period default probability.

We use maximum likelihood estimator to estimate the default probability $\hat{p}$ for each obligor. Using $\hat{p}$ we will estimate the probability that more than q% of the obligors will in default for q running from 0.1% to 10% for the next time period.

R code for Uniform Bernoulli Model Simulation

```
sam1<-replicate(20,sample(c(0,1),size=1000,replace = TRUE,prob=c(0.95,0.05)))

phat1<-length(sam1[sam1==1])/length(sam1)

p1<-1-pbinom(0.05*1000,1000,phat1)

prob<-c()

for(i in 1:100){

  prob[i]<-1-pbinom(i,N,phat1)

}

xrange <-c(0,0.1)

```

```
yrange <-c(0,1)

percentage<-c(seq(0,0.1,length=100))

probabilty<-prob

plot(yrange~xrange, type="n", xlab="percentage of obligors in default",
     ylab="probability" )

lines(probabilty~percentage, lwd=1.5,type="o",
       col="blue")
```



The picture shows that the probability that more than q% of the obligor is in default equals 1 if $q = 0$, as expected, and goes down to 0 as q goes to 100%. The steapest decay is around 0.05(5%), which is the default probability for each obligor.

5.2.2 Simulation of the Factor Model

The simulation procedure of the Factor Model are the following.

1. Data Production
We take the number of obligors $N = 1000$, time duration $T = 20$(years), and we let the sensitivity coefficient R vary in the range(0.2,0.4,0.6,0.8). If the log of obligor's asset value satisfies $r_i(t) < -c$, then it means $obligor_i$ in default. We choose again the default probability p equal to 0.05. So $-c = Q(0.05)$, where Q is the quantile function of standard normal distribution. $\Phi(t)$ is the composite factor of obligors, and $\Phi(t) \sim N(0,1)$. We generate 20 $N(0,1)$ random variables for $\Phi(t)$ each corresponding to one of the 20 years. Moreover we generate $20 \times N$ $N(0,(1-R^2))$ independent random variables for $\varepsilon_i(t)$. With these values we compute for each of the 20 years the values

$$r_i(t) = R\Phi(t) + \varepsilon_i(t), \qquad 1 = 1,...N.$$

For each of the 20 years we compute L_i,$i = 1,...N$, by $L_i = 1$ if $r_i < -c$ and $L_i = 0$ otherwise.

2. Estimate the next time period default probability.
We use maximum likelihood estimator to estimate the default probability $\hat{p}$. We estimate $\hat{V}$, $\hat{c}$, and $\hat{R}$ by means of the formulas in section 5.12, averaged over the 20 years. With these parameters we compute the probability that more than $(0\%, 10\%)$ of the obligors will be in default in the next time period. We do the whole procedure for each of the sensitivity coefficients R.

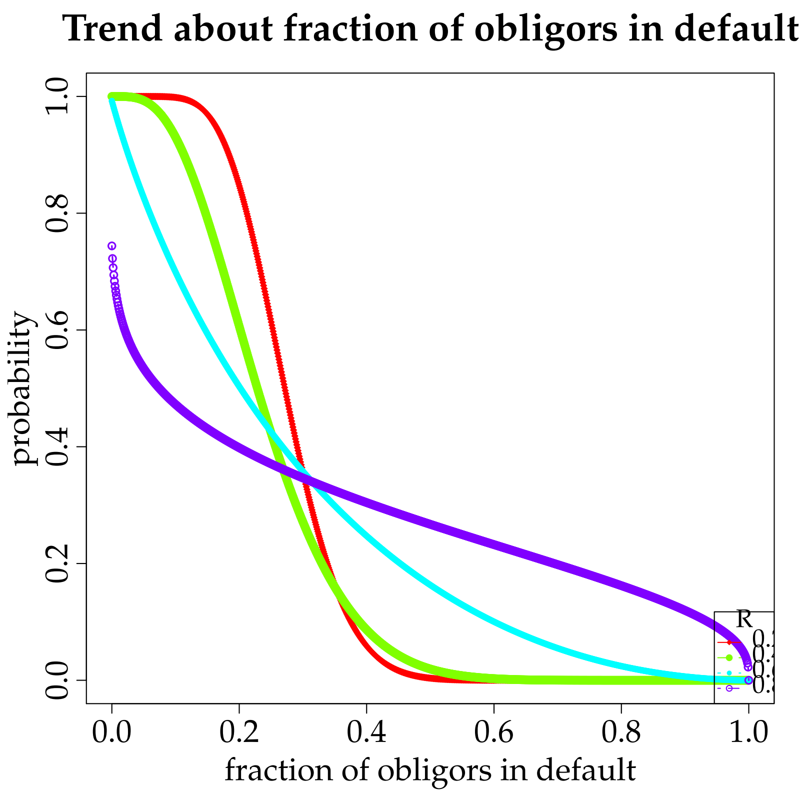
R code for Factor Model Simulation

```
R<-0.2

f<-function(R){

N<-1000

c<--qnorm(0.3)

theta<-rnorm(20,mean=0,sd=1)

epsilon<-replicate(20,rnorm(1000,0,1-R^2))
```

```
r<-epsilon
for(i in 1:20){
  r[,i]<- R*theta[i]+epsilon[,i]
}

sam2<- apply(r,2,function(x) as.numeric(x<=-c))

#calculate the calculate the probability that more than 10\%
#obligors will go into default directly from the data
default<-NULL
for(i in 1:20){

  default[i]<-sum(sam2[,i])

}

length(default[default>100])/length(default)

phat2<-length(sam2[sam2==1])/length(sam2)

chat<--qnorm(phat2)

V<-NULL

for(i in 1:20) {

  V[i]<-sum(sam2[,i])^2

  Vhat<-1/20*(sum(V))

}

a<-NULL

g<-function(a){
  integrate(function(x){
    (pnorm((-chat-a*x)/sqrt(1-a^2)))^2*(1/sqrt(2*pi))*exp((-x^2)/2)
  },-Inf,Inf)
}

h<-function(a){
  g(a)$value-(Vhat-N*phat2)/(N*(N-1))
```

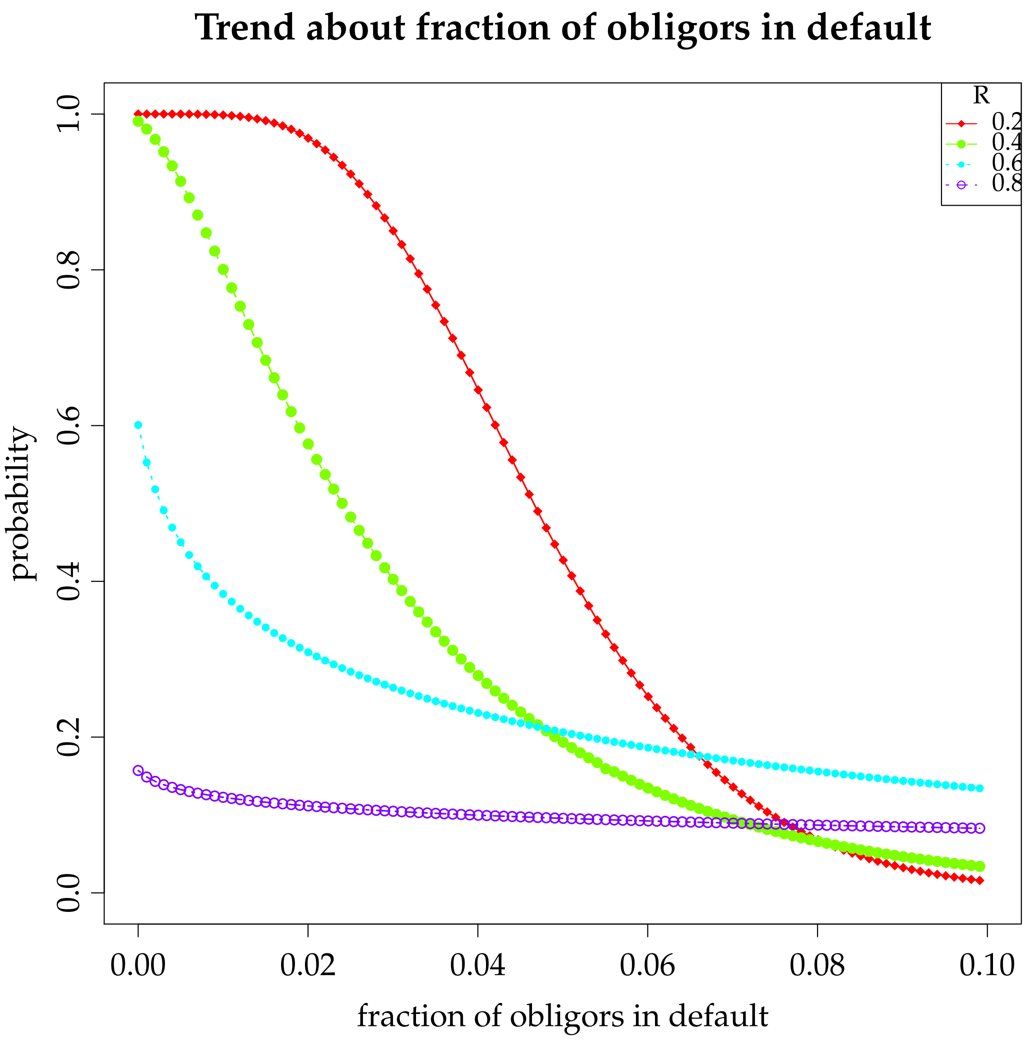
```
  }
Rhat<-uniroot(h,c(0,0.99999))$root
Rhat
}
result<-replicate(1000,f(R))
Rhat<-mean(result)
prob0.2<-c()
for(i in 1:1000){
  e<-function(s){
    pbinom(i,N,pnorm((-chat-Rhat$root*s)/sqrt(1-Rhat$root^2)))*dnorm(s)
  }
  prob0.2[i]<1-integrate(e,-Inf,Inf)$value
}

#Change R from 0.2 to 0.4,0.6,0.8, run pervious code again
#we can get the corresponding default probability (prob0.4,prob0.6,prob0.8).

Default<-data.frame(R=rep(c(0.2,0.4,0.6,0.8),each=100),
Probabilty=c(prob0.2,prob0.4,prob0.6,prob0.8),
percentage=c(seq(0,0.1,length=100),seq(0,0.1,length=100),
seq(0,0.1,length=100),seq(0,0.1,length=100)))

# Create Line Chart

# convert factor to numeric for convenience
Default$R <- as.numeric(Default$R)

nR <- max(Default$R)

# get the range for the x and y axis
```

```
xrange <- c(0,0.1)

yrange <- c(0,1)

# set up the plot
plot(yrange~xrange, type="n", xlab="percentage of obligors in default",
     ylab="probability" )

colors <- rainbow(4)

linetype <- c(1:4)

plotchar <- seq(18,18+4,1)

# add lines
for (i in 1:4) {
  R <- subset(Default, R==as.numeric(levels(as.factor(Default$R)))[i])
  lines(R$Probabilty~R$percentage, lwd=1.5,type="o",
        lty=linetype[i], col=colors[i], pch=plotchar[i])
}

# add a title and subtitle
title("Trend about percentage of obligors in default ")

# add a legend
legend("bottomleft",xrange[1], c(0.2,0.4,0.6,0.8), cex=0.8, col=colors,
       pch=plotchar, lty=linetype, title="R")
```



With the default probability 0.3. We see from the picture that the curves decrease from 1 to 0, as expected. The curves become less and less steep if the factor R increases. Since R is the coefficient describing the dependence of the joint factor Φ, it is expected that the default correlations get higher if R increases. We see in the picture that this indeed happens. In terms of credit risk management, this means that the probability of a large loss gets considerably higher for higher values of R, where the default probability is still the same value.
In case there were no correlation in the dataset (i.e. $R = 0$), the expected percentage of obligors in default in case the default probability is 0.3 would be 0.3. Moreover, in that case the probability that more than a fraction q of the obligors are in default would be very small if $q > 0.3$ and almost one if $q < 0.3$. For small correlation we see that the situation is similar, but the larger the correlation the larger the spread. More precisely, the probability that more than a fraction q of the obligors is indefault increases with increasing R if $q > 0.3$ and decrease with R if $q < 0.3$. In other words, the tail events get a higher probability if R is big. If $R = 0.6$ or $R = 0.8$ there is a significant probability that more than 80% of the obligors will be in default.

Moreover in the R code above we calculated the probability that more than 10% of obligors will go into default directly from the data. The results are 1, 1, 0.65, 0.5 corresponding with R 0.2, 0.4, 0.6, 0.8. The results are almost the same as computing method based on $\hat{p}$ and $\hat{R}$.



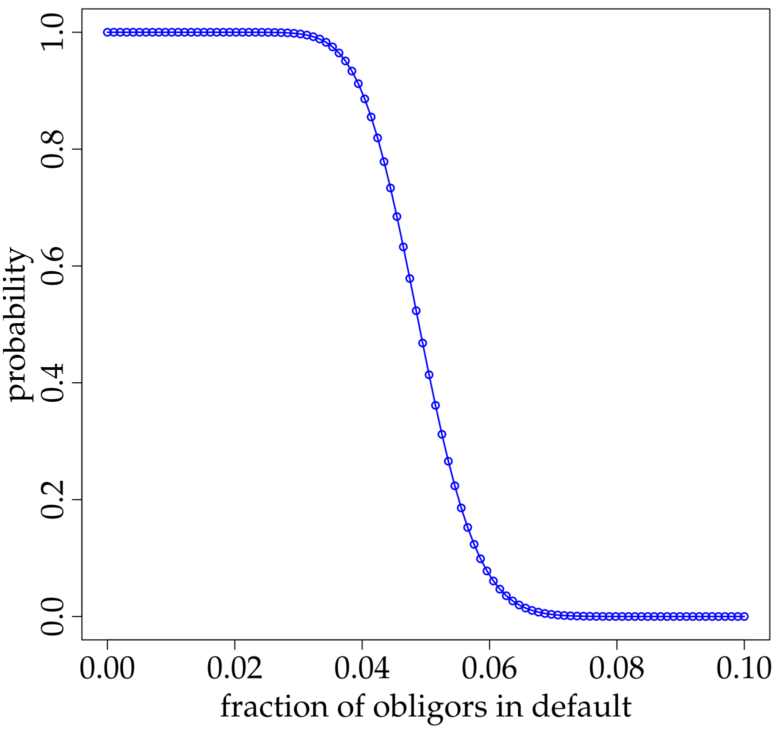
Lower default probability(0.05): More or less the same situation, but the decay of the curve now take place for lower percentage. This is as expected, since for a lower default rate the probability that more than 10% of the obligors are in default will be much lower.

5.2.3 Using factor model method on Bernoulli model dataset

```
N<-1000
```

```
sam1<-replicate(20,sample(c(0,1),size=1000,replace = TRUE,prob=c(0.95,0.05)))

phat1<-length(sam1[sam1==1])/length(sam1)

chat<--qnorm(phat1)

V<-NULL

for(i in 1:20) {

  V[i]<-sum(sam1[,i])^2

  Vhat<-1/20*(sum(V))
}

a<-NULL

g<-function(a){
  integrate(function(x){
    (pnorm((-chat-a*x)/sqrt(1-a^2)))^2*(1/sqrt(2*pi))*exp((-x^2)/2)
  },-Inf,Inf)
}

h<-function(a){
  g(a)$value-(Vhat-N*phat1)/(N*(N-1))
}

Rhat<-uniroot(h,c(0,0.99999))

prob<-c()

for(i in 1:100){

  e<-function(s){

    pbinom(i,N,pnorm((-chat-Rhat$root*s)/sqrt(1-Rhat$root^2)))*dnorm(s)
  }

  prob[i]<-1-integrate(e,-Inf,Inf)$value
}

xrange <- c(0,0.1)
```

```
yrange <- c(0,1)

percentage<-c(seq(0,0.1,length=100))

probabilty<-prob

plot(yrange~xrange, type="n", xlab="percentage of obligors in default",
     ylab="probability" )

lines(probabilty~percentage, lwd=1.5,type="o",
      col="blue")
```



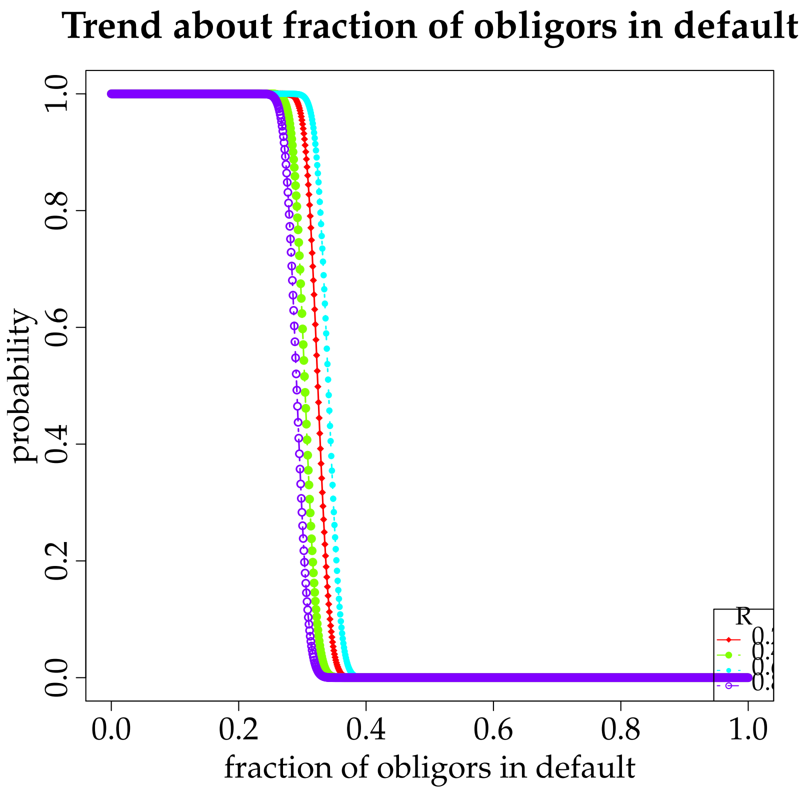
The picture shows that curve is almost the same as using the Bernoulli model method on Bernoulli dataset. This is not a surprise, since the Bernoulli model can be seen as a special case of the factor model with sensitivity coefficient R equal to 0, that is, without correlations. The factor model estimate will find an estimated coefficient $\hat{R}$ which is almost 0 since there are no correlations in the dataset. The estimated curve will then be almost the same as for the Bernoulli model. Hence using the factor model if the dataset is uncorrelated is safe.

5.2.4 Using Bernoulli model method on factor model dataset

```

N<-1000

R<-0.2

c<--qnorm(0.3)

theta<-rnorm(20,mean=0,sd=1)

epsilon<-replicate(20,rnorm(1000,0,1-R^2))

r<-epsilon

for(i in 1:20){

  r[,i]<- R*theta[i]+epsilon[,i]
}

sam2<- apply(r,2,function(x) as.numeric(x<=-c))

default<-NULL

for(i in 1:20){

  default[i]<-sum(sam2[,i])

}

length(default[default>300])/length(default)

phat2<-length(sam2[sam2==1])/length(sam2)

p2<-1-pbinom(0.05*1000,1000,phat2)

prob0.2<-c()

for(i in 1:1000){

  prob0.2[i]<-1-pbinom(i,N,phat2)
```

```

}

#Change R from 0.2 to 0.4,0.6,0.8, run pervious code again
#we can get the corresponding default probability (prob0.4,prob0.6,prob0.8).

Default<-data.frame(R=rep(c(0.2,0.4,0.6,0.8),each=100),
Probabilty=c(prob0.2,prob0.4,prob0.6,prob0.8),
percentage=c(seq(0,0.1,length=100),seq(0,0.1,length=100),
seq(0,0.1,length=100),seq(0,0.1,length=100)))

# Create Line Chart

# convert factor to numeric for convenience

Default$R <- as.numeric(Default$R)

nR <- max(Default$R)

# get the range for the x and y axis
xrange <- c(0,1)
yrange <- c(0,1)

# set up the plot

plot(yrange~xrange, type="n", xlab="percentage of obligors in default",
     ylab="probability" )

colors <- rainbow(4)

linetype <- c(1:4)

plotchar <- seq(18,18+4,1)

# add lines
for (i in 1:4) {
  R <- subset(Default, R==as.numeric(levels(as.factor(Default$R)))[i])
  lines(R$Probabilty~R$percentage, lwd=1.5,type="o",
          lty=linetype[i], col=colors[i], pch=plotchar[i])
}

# add a title and subtitle
title("Trend about percentage of obligors in default ")
```

```

# add a legend
legend("bottomright",xrange[1], c(0.2,0.4,0.6,0.8), cex=0.8, col=colors,
        pch=plotchar, lty=linetype, title="R")

lines(probability~percentage, lwd=1.5,type="o",
      col="blue")
```



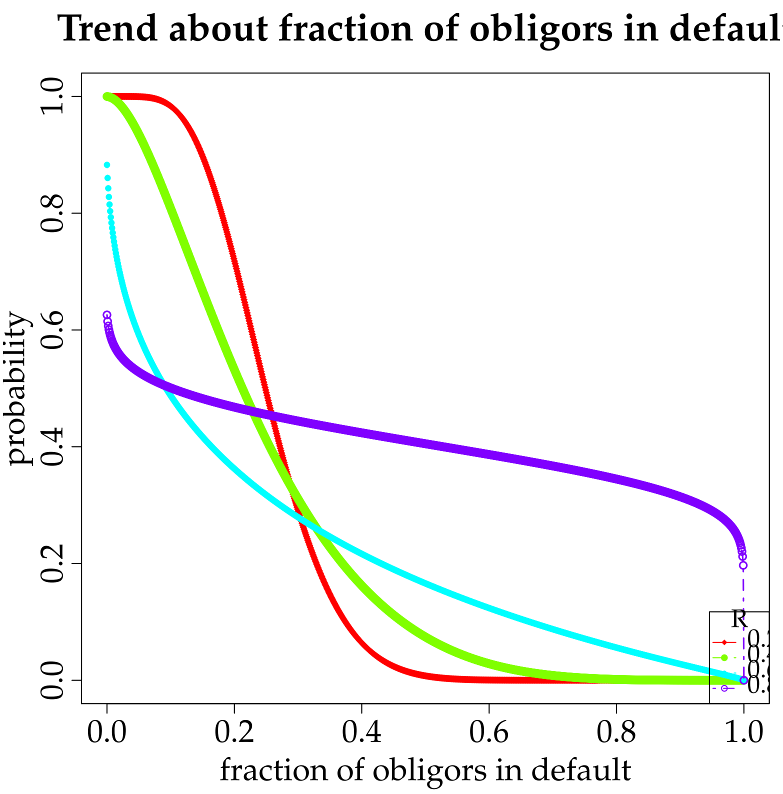
The picture shows that the prediction with the Bernoulli model are very different from those with the factor model. The likeness of a fraction of defaults that differs from the default probability is estimated to be very small. This estimate is highly inaccurate. Indeed, one can verify directly from the data that the actual occurrences of time periods with more than a fraction q (with $q > 0.3$) in default is 0.4 which is much more frequent than the almost zero prediction. This shows that ignoring correlation may lead to a severe underestimation of default risks in the portfolio.

5.2.5 Make a new dataset by two factor model

```
N<-1000

R1<-0.2

R2<-0.3

c<--qnorm(0.3)

theta1<-rnorm(20,mean=0,sd=1)

theta2<-rnorm(20,mean=0,sd=1)

epsilon<-replicate(20,rnorm(1000,0,1-R1^2-R2^2))

r<-epsilon

for(i in 1:20){

   r[,i]<- R1*theta1[i]+R2*theta2[i]+epsilon[,i]
}

sam3<- apply(r,2,function(x) as.numeric(x<=-c))

default<-NULL

for(i in 1:20){

   default[i]<-sum(sam3[,i])

}

length(default[default>400])/length(default)

phat2<-length(sam2[sam2==1])/length(sam2)

phat3<-length(sam3[sam3==1])/length(sam3)

chat<--qnorm(phat3)

V<-NULL
for(i in 1:20) {
```

```
  V[i]<-sum(sam2[,i])^2

  Vhat<-1/20*(sum(V))
}

a<-NULL

g<-function(a){
  integrate(function(x){
    (pnorm((-chat-a*x)/sqrt(1-a^2)))^2*(1/sqrt(2*pi))*exp((-x^2)/2)
  },-Inf,Inf)

}

h<-function(a){
  g(a)$value-(Vhat-N*phat2)/(N*(N-1))}

Rhat<-uniroot(h,c(0,0.99999))

prob0.2<-c()

for(i in 1:1000){

  e<-function(s){

    pbinom(i,N,pnorm((-chat-Rhat$root*s)/sqrt(1-Rhat$root^2)))*dnorm(s)
  }

  prob0.2[i]<-1-integrate(e,-Inf,Inf)$value
}

#Change R1 from 0.2 to 0.4,0.6,0.8, run pervious code again
#we can get the corresponding default probability (prob0.4,prob0.6,prob0.8).

Default<-data.frame(R=rep(c(0.2,0.4,0.6,0.8),each=100),
Probabilty=c(prob0.2,prob0.4,prob0.6,prob0.8),
percentage=c(seq(0,0.1,length=100),seq(0,0.1,length=100),
seq(0,0.1,length=100),seq(0,0.1,length=100)))

# Create Line Chart

# convert factor to numeric for convenience
Default$R <- as.numeric(Default$R)
```

```
nR <- max(Default$R)

# get the range for the x and y axis
xrange <- c(0,1)

yrange <- c(0,1)

# set up the plot
plot(yrange~xrange, type="n", xlab="percentage of obligors in default",
     ylab="probability" )

colors <- rainbow(4)

linetype <- c(1:4)

plotchar <- seq(18,18+4,1)

# add lines
for (i in 1:4) {
  R <- subset(Default, R==as.numeric(levels(as.factor(Default$R)))[i])
  lines(R$Probabilty~R$percentage, lwd=1.5,type="o",
        lty=linetype[i], col=colors[i], pch=plotchar[i])
}

# add a title and subtitle
title("Trend about percentage of obligors in default")

# add a legend
legend("bottomright",xrange[1], c(0.2,0.4,0.6,0.8), cex=0.8, col=colors,
       pch=plotchar, lty=linetype, title="R")
```



The picture shows that the one-factor model applied to the two-factor data gives a prediction of default curves that look similar to those of the one-factor dataset. The spreads are a bit wider than for the curves of section 5.2.2, which is due to the additional market factors term which introduces additional correlation.
If we compare the estimated probabilities of more than a fraction 0.4 in default (with 0.4 larger than the default probability) with the counted fraction of the data set which has results are 0, 0.15, 0.2, 0.45 corresponding with R 0.2, 0.4, 0.6, 0.8. We see that the predictions are quite good. This suggest that the correlations are much more important than the precise way in which they are induced. The one-factor model seems to capture the risks due to correlation quite well.

References

[1] Anthony, S., Marcia, C. (2010).Financial Institutions Management: A Risk Management Approach, McGraw-Hill/Irwin; 7 edition.

[2] Bluhm, C., Overbeck, L., Wagner, C. (2002). An introduction to credit risk modeling. CRC Press.

[3] Chan-juan,C and Hai-bo Z. "Toward the Application of CPV Model in the Calculation of Loan Default Probability [J]." Modern Economic Science 5 (2009): 006.

[4] Du, Y., Suo, W. (2003). Assessing Credit Quality from Equity Markets: Is Structural Model a Better Approach?. Available at SSRN 470701.

[5] Faraway (2006). Extending the linear model with R. Generalized linear, mixed effects and nonparametric regression models. Chapman Hall/CRC

[6] Gordy, M. B. (2003). A risk-factor model foundation for ratings-based bank capital rules. Journal of financial intermediation, 12(3), 199-232.

[7] Gordy, M. B. (1998). A comparative anatomy of credit risk models.

[8] Gelman, A., Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

[9] Jiang, L. (2005). Mathematical modeling and methods of option pricing. Singapore: World Scientific.

[10] Junxian L., Maureen Olsson L(2009). "CREDIT RISK MANAGEMENT OF THE CHINESE BANKS BASED ON THE KMV MODEL " Master's Degree Thesis.

[11] Korablev, I., Dwyer, D. (2007). Power and level validation of Moody's KMV EDF credit measures in North America, Europe, and Asia, Technical Paper, Moody's KMV Company. Page 7.

[12] Kupiec, P. H. (2008). A generalized single common factor model of portfolio credit risk. The Journal of Derivatives, 15(3), 25-40.

[13] Peter, C., Ahmet, K. (2003). MODELING DEFAULT RISK . Technical Paper, Moody's KMV Company.

[14] Thorp, J. (2003). Change of seasonal adjustment method to X-12-ARIMA. Monetary and Financial Statistics, 4-8.

[15] Website of National Bureau Of Statistics Of China,(http://www.stats.gov.cn/).

[16] Website of China Banking Regulatory Commission,(http://www.cbrc.gov.cn/index.html).

[17] Website of Centraal Bureau voor de Statistiek,(http://www.cbs.nl/en-GB/menu/home/default.htm).

[18] Website of De centrale bank van Nederland,(http://www.dnb.nl/home/index.jsp).

[19] Website of the Netease Finance,(http://quotes.money.163.com).

[20] Website of YAHOO Finance,(http://finance.yahoo.com/).

[21] Xiaohong,C., Xiaoding, W, and Desheng, Wu. "Credit risk measurement and early warning of SMEs: An empirical study of listed SMEs in China." Decision Support Systems 49, no. 3 (2010): 301-310.