[Inner endpaper]

HEC MONTRÉAL

Modelling CDS Market-implied Recovery Rates with the Beta Distribution

par

Hon Cheung Hui

Pascal François

HEC Montréal

Directeur/Directrice de recherche

Sciences de la gestion
(Spécialisation Financial Engineering)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

May 2021

Résumé

Cette thèse extrait les taux de recouvrement implicites du marché des CDS en utilisant la technique développée par Das et Hanouna (2009). Ce travail examine la modélisation de ces taux de recouvrement prospectifs à partir de la distribution bêta. Cette distribution correspond bien avec la distribution empirique des taux de recouvrement. Pour tenir compte des variations dans le temps des taux de recouvrement, une estimation VAR est mise en œuvre pour l'analyse des séries chronologiques. Divers tests comme le test de racine unitaire et BIC sont adaptés avec l'estimation VAR pour la sélection du retard. Les résultats pour les prévisions en échantillon et pseudo hors échantillon sont présentés. Même s'il y a des différences entre industries, la RMSE pour la prévision jusqu'à 6 périodes à venir demeure faible.

Mots clés : Taux de recouvrement implicite, risque de crédit, distribution bêta

Méthodes de recherche : Modèle binomial avec défaut immédiat, VAR

Abstract

The thesis extracts market-implied recovery rates from CDS using the technique developed by Das and Hanouna (2009). This work examines the modelling of these forward-looking recovery rates with the beta distribution. This distribution fits well with the empirical distribution of recovery rates. To accommodate the time-varying feature of the recovery rates, a VAR estimation is implemented for the time series analysis. Various tests like unit root test and BIC are adapted along with VAR estimation for lag selection. The results for both in-sample and pseudo out-of-sample forecasts are presented. Despite variations across industries, the RSME for forecasting up to 6 periods ahead is still small.

Keywords: Implied recovery rate, Credit risk, Beta distribution

Research methods: Binomial model with jump to default, VAR

Table of Contents

List of tables

List of figures

List of abbreviations and acronyms

RR	Recovery Rate
LGD	Loss given default
PD	Probability of Default
CDS	Credit Default Swaps
VAR	Vector Auto-Regression
NAICS	North American Industry Classification System
SIC	Standard Industrial Classification
PDF	Probability Density Function
RMSE	Root Mean Square Error
BIC	Bayesian information criterion
ADF	Augmented Dickey-Fuller
LB	Ljung–Box

Preface

Risk management is getting more and more attention from regulators around the globe. The interest in credit risk and derivatives have driven me to study CDS with risk management application. The forward-looking and non-constant assumption for recovery rate is better for risk control purposes. I hope this paper could shed some light on how we could model the recovery rate from market implied information.

Acknowledgements

I would like to express my gratitude to my research director, Pascal François, for his support and patience throughout this thesis. His advice and feedback have helped me a lot. His support is crucial to this thesis. Without him, it is impossible for me to finish this thesis. I really appreciate the opportunities to work under him and thank for his support.

Also, I would like to thank my beloved families and friends who have motivated me and support me during my studies.

Introduction

In 2001, the Basel committee introduced internal rating-based approach for financial institutions to calculate its capital requirements (Bank for International Settlements 2001). Various factors are required for the capital calculation like Probability of Default (PD), Loss Given Default (LGD), which is one minus Recovery Rate (RR) etc. For financial institutions developing its advanced internal ratings based (AIRB) approach to calculate its capital requirements, Basel II and III provide incentives for them in terms of regulatory requirement.

In this paper, among those risk factors, the recovery rate will be our focus. Apart from regulatory aspects, modelling RR is also critical for credit derivatives pricing. As stated in Merton (1974) work, the credit spread can be calculated by the product of PD and LGD under the risk neutral measure. Since recovery rate is entangled with probability of default in credit spread, people usually assume it is constant in order to determine the relationship of credit spread and probability of default. For example, Giesecke et al. (2011) assume a 50% recovery rate. However, various research are extracting the recovery rate using different methods recently.

In this paper, I will use the technique from Das and Hanouna (2009) and François and Jiang (2019) to extract the market implied recovery rate. Then, beta distribution will be used for modelling and forecasting.

For fitting the empirical data, beta distribution estimations show a good fit. For time series analysis of beta distribution using VAR (p), both in sample fitting and pseudo out-of-sample forecasts have delivered promising result. Among different maturities examined, 5-year one provided the best result. Its 6-month forecast RSME to mean ratio is just 3.6%.

Chapter 1
Literature review

1.1 Historical recovery rate

Numerous articles have examined the determinants of historical recovery rate. They can be grouped into several categories. The debt structures like seniority, collateralization are not included in this review since Credit Default Swaps (CDS) contracts are the focus of our paper. When CDS contract is triggered, it is settled with buyer receiving protection payout. It is triggered by any occurrence of default or restructuring event on a company Therefore, CDS is terminated by the single triggering event on the reference company regardless of how senior or even overcollateralized its other debts are.

Firm-specific factors

It is not surprising that firm-specific factors play a role in recovery rate. However, their availability is somewhat limited. In the sample of Varma and Cantor (2005), among 2,000 corporate issuers that defaulted between 1983 and 2003 according to Moody's proprietary default database, only around 1,100 issuers had complete financial information before default. The ratio is roughly just above half, 55%. Varma and Cantor (2005) conducted both univariate and multivariate analyses on the following factors: Leverage, Market-to-book ratio, Tangibility, Earnings, Stock returns etc. Leverage is measured by the ratio of all publicly traded debt to total assets. It is found that high leverage ratio firms have just 5% lower in recovery rate than those with low ratio, but it is statistically significant. Market-to-book ratio, calculated as the sum of the market value of equity plus the book value of debt divided by the book value of assets, has a positive correlation factor with recovery rate, i.e., higher this ratio is, higher the recovery rate. Tangibility is measured by the ratio of hard assets to total assets. Higher tangibility firms enjoy around 25% more in recovery rate than firms with lower tangibility. Earnings, calculated by EBITDA-to-sales, surprisingly did not show a strong evidence on its impact on recovery rate. Stock returns, like earnings, is not showing neither economically nor statistically significance on recovery rate. The authors also found that firm size (total assets) is not a determinant of recovery rate significantly.

In the previous paper, earnings are not an important factor for recovery rate. In fact, Acharya et al. (2007) do not study earnings but the profit margin. Using OLS estimates of regression, they find that profit margin is positively correlated with recovery rate. Similar to Varma and Cantor (2005), the firm size cannot be a determinant of recovery rate statistically.

Industry-specific factors

Altman and Kishore (1996) use three-digit Standard Industrial Classification codes to classify the data into 18 industries with 696 observation in total. In the period, from 1971 to 1995, they find that public utilities sector has the highest recovery rate (RR) with 70.47 (per 100 face value) on average while lodging, hospitals, and nursing facilities have the lowest, 26.49. Even taking the seniority into account, the differences between sectors are statistically significant. Besides, they find that original ratings, size of issuance and time to default from issuance date have no impact on the RR. This clearly demonstrates that industry-specific factors do play a role on the RR.

Through which channel do the industry-specific factors affect the recovery rate? One intuitive answer is the macroeconomic environment. When the economy is in recession overall, it is inevitable that all industries are affected. The remaining value of an asset, recoverable amount, of the defaulted firm is decreased as there is an economic downturn. Shleifer and Vishny (1992) make the following important distinction. Some assets like commercial land, can be operated for various usages. Therefore, upon default of the asset holder, this kind of asset can be bought from both within and outside the industry. It merely depends on the overall economic cycle. However, for most assets, like aircraft, oil rig, machinery for a particular manufacturing process etc,, a sale to an industry outsider is not easy. If the default is caused by idiosyncratic shocks like mismanagement, these assets could be easily acquired by peers. If the industry is in distress while the economy is not in recession, other companies within the sector are likely to be in trouble. In this case, only an outsider firm may acquire the asset, and the resale value is likely to be priced below its best use value. As a result, distress in industry causes fire-sales which could lower recovery rates.

As mentioned above, macroeconomic recession could also cause distress in industry and impact on recovery rate. To further study the industry factor, this economic downturn should be disentangled from industry distress factor. Acharya et al. (2007) disentangle the distress factor from the economic downturn factor. Their empirical study examines defaulted securities in the United States over the period 1982–1999. First, they use 3-digit SIC code for classification of industry. After that, they use firms stock price as a proxy for an indicator whether the industry is in distress or not. If median stock return for this industry is less than or equal to 30%, then this industry is classified in distress. After that, they measure industry's asset-specificity as the median book value of the industry's machinery and equipment divided by the book value of total assets. When the industry is not in distress, there is no correlation between recoveries and asset specificity. But when it is in distress, the correlation is 0.76 and statistically significant. Also, the magnitude is 11.11% less compared to non distress period. Besides, firms take 2.16 years in bankruptcy process during industry distress period while they spend 1.37 years when there is no distress in the industry. Acharya et al. (2007) further show that an industry in distress will greatly reduce the recovery rate even when the overall economy is not recession.

Gupton and Stein (2002 and 2005) also find that RR across industries have different average values and different distribution shapes. In addition to three other factors, Seniority, Macro Economic and Firm Specific, industry accounts for around 20 % to predict RR. Furthermore, the industry RR are not static over time. For example, the distribution of RR for the telephone sector shifted towards lower values after 1998. This is probably due to the obsolescence cycle the industry faced at that time as new technology like wireless communication emerged.

Macroeconomic-specific factors

Khieu et al. (2012) use GDP growth as a proxy for macroeconomic conditions and find it is positively and significantly correlated with recovery rate. Tobback et al. (2014) obtain a similar finding. In addition, Khieu et al. (2012) find that unemployment rate is also a factor with negative correlation. In Sabato and Schmid (2008), although GDP growth and unemployment rates could help to predict the recovery rate changes, these indicators are lagging indicator of economic downturn. Therefore, the effect on recovery rate could also be lagged.

Nazemi at al. (2018) extensively examine 104 macroeconomic variables in predicting recovery rate using multi-factor support vector regressions. Among the top 20 most significant variables, 6 are interest rate related factors like different corporate bond yields and yield spreads, or Treasury 10-year bond yields. Also, 10 of these variables are stock market indicators like various index returns and volatilities.

Altman et al. (2005) explore the relationship between default rates and recovery rate. The economic intuition is clear. Consider the demand and supply of distressed securities in a recession, the default rate goes up. Therefore, the supply goes up, but the demand does not grow with the pace of supply, thereby causing the price of asset and recovery rate to drop. Besides, the macroeconomic factors like GDP growth drops in recession, this could increase default rate. At the same time, it lowers recovery rate as asset price drops too. When testing with a univariate model, GDP growth and default rate do have positive and negative correlation with recovery rate respectively. Surprisingly, when GDP growth is included in multivariate models, this macroeconomic variable is not just becoming insignificant but with negative sign too.

Shortfalls

First, to construct the analysis, availability of data is an issue. The data for modelling recovery rate must be wide enough in term of cross section i.e., including various firms and industries with different credit rating. At the same time, the timeframe should be sufficiently long to include full economic cycle as credit cycle do affect the recovery rate (Bruche and Gonzalez-Aguado, 2010). For firm-specific factors, it may be difficult to collect such data.

Second, historical recovery rate is an ex-post measurement, it is backward looking into the past observations. It is determined under the physical measure, not under the risk neutral measure. Therefore, it does not take any risk premium into account. It is important for risk management and pricing credit derivatives to include this risk premium.

1.2 Market-implied recovery rate

In equity options pricing, there are historical and implied volatility. The latter is extracted from the traded derivatives in the market. It represents a risk-neutral measure and includes market implied information. In credit default swap (CDS), with various models, the market-implied recovery rate could be extracted, this is critical for pricing and risk management purpose. However, this calculation is not as straightforward as options. The credit spread can be calculated as the product of PD and LGD under the risk neutral measure in reduced-form models. To make things more complicated, Altman et al. (2005) have shown that default rates and recovery rate are correlated. Since implied PD and LGD cannot be directly observed from the market, we need different techniques to decompose these two variables from observed market price of CDS.

Some papers assume constant recovery rate for extracting information from traded CDS like Lipton and Sepp (2009) or Finger et al. (2002). This assumption simplifies the pricing issue, but it may be impractical at the same time. The static recovery assumption (either constant or cross sectional) is far from reality. The recovery rate is not only affected by the cross section of underlying firms but also affected by time series factors like the credit cycle as shown in the preceding section. Therefore, a dynamic model is required to extract this non static market implied recovery rate.

In digital default swaps, the recovery rate is pre-determined. Therefore, the only variable in payout amount is PD. Berd (2005) shows that using this instrument and a standard conventional CDS, we can disentangle recovery rate from PD. However, this technique depends on the availability of digital default swaps (or recovery swaps), but it is only a small niche in the credit derivatives market. This could be a problem if there is no digital default swaps with same underlying reference.

Unal et al. (2003) use the adjusted relative spread between senior and junior debt. At the time of default, debtors’ default on all outstanding debt obligations regardless of their seniority. Therefore, the recovery rate can be isolated from the price.

In this paper, I follow Das and Hanouna (2009) for extracting recovery rate from market information using a reduced form model. The instruments they use are CDS and equity for market information. First, default probability is inferred from stock prices using a Cox et al. (1979) binomial model with jump to default. Then, the recovery rate is parametrized with the default probability to match the term structure of CDS spreads. This model only uses three parameters which makes it parsimonious. Besides, once the parameters are estimated, the whole risk neutral term structure can be identified.

1.3 Beta distribution for recovery rate

When it comes to modelling the distribution of recovery rates, it is known that the beta distribution can be a good candidate. It is bounded within [0,1]. It is parsimonious as it requires only two inputs (α and β). In fact, it is used by rating agencies like Moody's LOSSCALC (Gupton and Stein, 2002). The two parameters could be further modelled. Bruche and Gonzalez-Aguado (2010) model these two parameters with an exponential function of four factors like industry, credit cycle, seniority, and macroeconomic variables.

Instead of a simple beta distribution, researchers also use an inflated beta regression model which is a mixture of beta distribution and Bernoulli distribution (Bellotti and Crook, 2012) to accommodate the bimodal distribution. In addition to the beta distribution, some researchers even use non-parametric methods to model recovery rate (Chen et al., 2019)).

Qi and Zhao (2011) find that non-parametric methods like neural networks perform better in terms of model fit and predictive accuracy compared to parametric methods. Interestingly, they find that whether the model can generate bi-modal distribution may not be the major concern. It is always a trade-off between tractability and accuracy. More complex methods can have a better forecasting ability but could be difficult to implement and interpret, or even suffer from a "black box" effect like neural network.

Chapter 2
Data and Methodology

The methodology section has three parts. The first part is devoted to extract market-implied recovery rates for each individual firm using the technique in Das and Hanouna (2009).In the second part, the beta distribution is estimated in the cross-section of firms for each month.

The last part consists in modelling the alpha and the beta in the time series using a Vector AutoRegression approach (VAR). Once the parameters are estimated, both in-sample and out-of-sample forecasts will be conducted.

2.1 Market-implied recovery rate from CDS

From Datastream, I collect monthly CDS spread term structures and the associated stock prices (denominated in U.S. dollars) between January 2005 and December 2019. In total, there are 180 months of data. The mid-spread, i.e., the average of bid and ask prices, is used for fitting. The data comprises 7 maturities from 1 to 5 years, 7 and 10 years. There are 351 firms included in the sample.

Also, discounted factors are required to retrieve the implied recovery rate. To begin with, I download monthly Treasury yields from 1 month to 10 years, 9 maturities in total, from Fed St-Louis website. Then, Nelson-Siegel curve fitting is performed to retrieve the required spot rates and forward rates for each dates and maturities.

Here is a brief description of the technique (please refer to Das and Hanouna, 2009, for more details.)

Using a binomial tree for the stochastic stock price as in Cox et al. (1979) with a 'jump-to-default' state, the term $\lambda[i,j]$ denotes the probability of a jump to default at node i at time j.

$[q,1-q]$ are the probabilities of the stock going up and down next period, respectively. If it is going up, the up factor is $u = \exp(\sigma\sqrt{h})$ where h is the time step of each node and σ

is the stock volatility. If it is going down, the down factor is d = exp(- $\sigma\sqrt{h}$). Let R = exp (fh) denote the compounding factor, where f is risk-free rate. The stock price $S[i,j]$ in each node can be computed as

$$S[i,j+1] = \begin{cases} S[i,j]\,u, & probability\; q(1-\lambda[\mathrm{i,j}]) \\ S[i,j]d, & probability\; (1-q)(1-\lambda[\mathrm{i,j}]) \\ 0, & probability\; \lambda[\mathrm{i,j}] \end{cases}$$

The risk neutral probability of going up is

$$q = \frac{\frac{R}{1-\lambda[i,j]} - d}{u-d}$$

For the default probability, $\lambda[i,j]$ and the recovery rate, $\psi[i,j]$, we have

$$\lambda[i,j] = 1 - exp(-S[i,j]^{-b}h) \tag{1}$$

$$\psi[i,j] = N(\,a_0 + a_1\,\lambda[i,j])$$

where $N(.)$ is the cumulative normal distribution function.

This model thus consists of four parameters (a_0, a_1, b, σ).

Let $\pi_j\;(a_0, a_1, b, \sigma)$ denote the premium of the CDS while π_j^{mkt} is the observed market premium of the CDS at time j. For σ, instead of using historical volatility, I follow François and Jiang (2019) and include it in the optimization. The CDS premium is given by

$$\pi_j\;(a_0, a_1, b, \sigma) = \frac{\sum_{u=1}^{j}(\prod_{k=1}^{u-1}(1-\lambda^k)\lambda^u(1-\psi^u)B(T_u)}{\sum_{u=1}^{j}(\prod_{k=1}^{u}(1-\lambda^k)B(T_u)}$$

where $B\,(t)$ is the discount factor from time 0 up to time t.

Parameters are obtained by fitting the observed CDS premiums with the following optimization program:

$$\min_{a_0,a_1,b,\sigma} \sum_{j=1}^{7} (\pi_j\,(a_0, a_1, b, \sigma) - \pi_j^{mkt})^2$$

Once the four parameters (a_0, a_1, b, σ) are estimated, the recovery rate for each month and its term structure can be computed by substituting the estimated parameters into recovery rate equation (1).

I focus on the 1-, 5- and 10-year maturities to span the common durations of corporate bonds.

2.2 Descriptive statistics

Table 1 shows the summary statistics for CDS spreads data. There are 351 firms across 15 years period, from January 2005 to December 2019. Each term structure consists of 7 maturities, from 1 year to 5 years, 7 years and 10 years. There are 422,153 observed spreads in total.

CDS maturity	**Maximum**	**Minimum**	**Mean**	**Std. Dev.**	**Skewness**
1	34,249.11	0.15	118.89	708.22	25.46
2	24,119.88	0.70	132.76	532.70	20.45
3	19,081.45	1.20	151.06	462.63	16.82
4	16,269.79	2.30	171.12	429.35	14.55
5	14,627.00	3.60	188.51	406.59	13.23
7	14,591.50	4.60	206.64	376.64	12.15
10	14,580.00	5.66	216.62	349.97	11.56

Table 1 Summary statistics for CDS spreads in basis points.

From Table 1, upward sloping CDS spread term structure is observed. From average of 119 bps in 1 year to 217 bps in 10 years. The volatility is high, e.g., 708bps for 1-year CDS with mean spread 119. This is probably due to the fact that I included the recession period, 2008 financial crisis in the sample period. The skewness is positive across maturity but shows monotonically decreasing trend with maturity.

CDS maturity	**Median**	**Mean**	**Std. Dev.**	**Skewness**
1	93.48%	89.02%	14.35%	-3.916
2	81.19%	77.86%	14.35%	-2.822
3	71.17%	68.69%	14.29%	-2.066
4	62.78%	61.07%	14.12%	-1.495
5	55.81%	54.69%	13.91%	-1.035
7	44.79%	44.63%	13.60%	-0.391
10	33.28%	34.04%	13.26%	0.200

Table 2 Summary statistics for recovery rate.

Table 2 reports the implied market recovery rate statistics. Table 2 clearly shows the constant recovery rate assumption is not consistent with the market implied recovery rate. The 1-year recovery rate is nearly double to 10-year one. The term structure is downward sloping with CDS maturity, which is consistent with François and Jiang (2019) and Das and Hanouna (2009) findings. This may be explained by the slow deterioration of the firm over the longer period. This implies the risk of asset value drop and a larger discount factor.

As mentioned in previous section 2.1, 1-year, 5-years and 10-years are the primary focus in this paper. Figure 1 shows the distribution of recovery rate for different maturities for the whole sample period. It demonstrates a unimodal distribution. Therefore, beta distribution could be used for modelling.

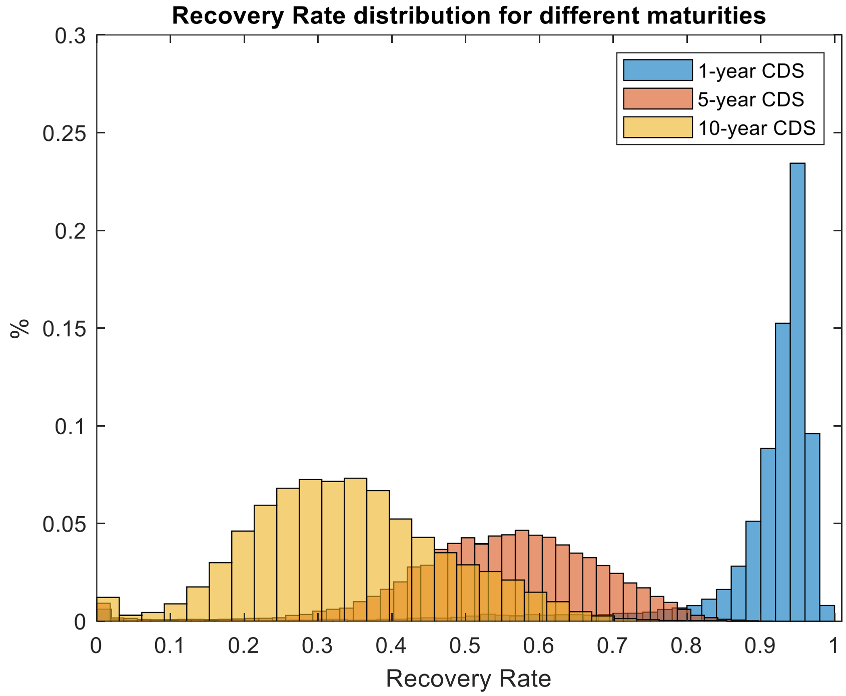


Figure 1 Market-implied Recovery Rate distribution for different maturities

By business cycle

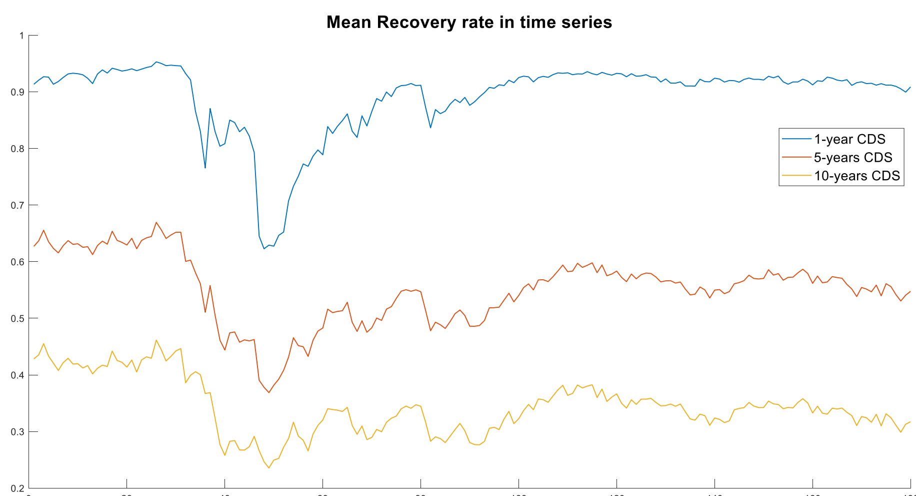


Figure 2 Mean Recovery Rate

Figure 2 reports the 1, 5 and 10-year CDS mean recovery rates in the sample period. It shows the longer maturity CDS has a more volatile recovery rate. The 5 and 10-year recovery rates did not recover to their own pre-crisis level even a long period of time after the crisis.

The samples are further divided into 3 subperiods: Before the recession, January 2005 - December 2007, 36 months, during the recession January 2008 - June 2009, 18 months and after the recession July 2009 - October 2014, 64 months.

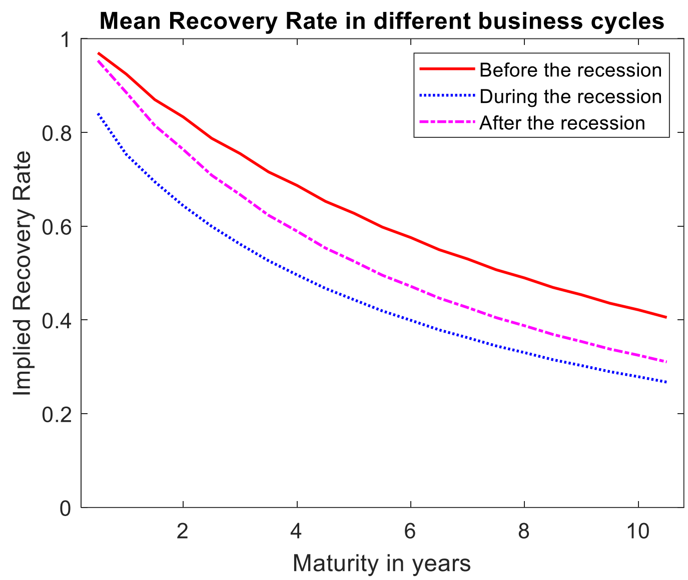


Figure 3 Mean Recovery rate in different business cycle

In Figure 3, from before to during the recession, the shock induced a parallel shift of the term structure. It is worth noting that after the recession, the shorter maturity RR rebounded much faster than the longer ones. This suggests that the market is still not confident about the long-term recovery of the economy. Besides, this further demonstrates the dynamic aspect of recovery rates.

By industry

As demonstrated in Altman and Kishore (1996), the recovery rate is affected by industry factors. Using firms' first 2-digit SIC (Standard Industrial Classification) code, I further classify their industries, ranging from Manufacturing to Finance and Insurance.

Since there are only limited samples in some industries, any industry with a number of firms lower than 20 is dropped in the industry analysis.

Industry	Number of firms
Manufacturing	134
Transportation, Communications, Electric, Gas, And Sanitary Services[1]	64
Finance, Insurance, And Real Estate[2]	57
Retail Trade	28
Mining	25
Services	24
Construction	9
Wholesale Trade	6
Public Administration	4

Table 3 Number of firms in industry

Therefore, I only include Manufacturing, Transportation, Finance, Retail, Mining and Services.

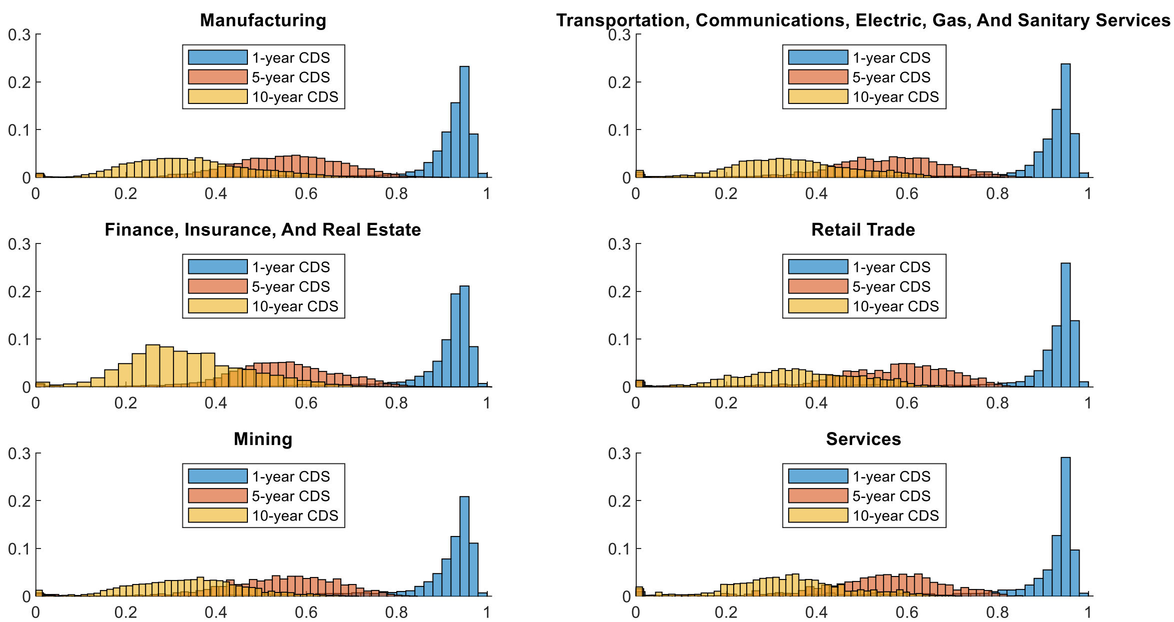


Figure 4 Recovery rate distribution by Industry

[1] Denoted as Transportation.

[2] Denoted as Finance.

Industry		Manufacturing	Transportation	Finance	Retail Trade	Mining	Services	Whole sample
Number		134	64	57	28	25	24	351
1-year CDS	**Median**	93.43%	93.60%	93.09%	94.23%	93.35%	93.93%	93.48%
	Mean	89.72%	88.97%	88.95%	89.23%	87.87%	86.34%	89.02%
	Std. Dev.	12.66%	14.83%	13.64%	15.95%	15.23%	18.83%	14.35%
	Skewness	-4.248	-3.890	-4.089	-3.889	-3.274	-2.794	-3.916
	***p*-value**	*0.0881*	*0.8554*	*0.7442*	*0.8163*	*0.1631*	*0.0001*	N/A
	Reject?	N	N	N	N	N	Y	N/A
5-year CDS	**Median**	55.84%	56.47%	54.43%	58.54%	55.13%	55.26%	55.81%
	Mean	55.02%	54.95%	54.05%	56.16%	53.49%	53.20%	54.69%
	Std. Dev.	13.38%	14.47%	13.15%	14.84%	14.10%	15.55%	13.91%
	Skewness	-0.879	-1.137	-0.884	-1.398	-1.094	-1.117	-1.035
	***p*-value**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	N/A
	Reject?	Y	Y	Y	Y	Y	Y	N/A
10-year CDS	**Median**	32.89%	33.70%	31.76%	35.53%	33.93%	33.54%	33.28%
	Mean	33.98%	34.51%	33.26%	35.86%	33.64%	33.70%	34.04%
	Std. Dev.	13.32%	13.53%	13.09%	13.59%	12.41%	13.29%	13.26%
	Skewness	0.331	0.103	0.440	-0.131	-0.059	-0.050	0.200
	***p*-value**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	N/A
	Reject?	Y	Y	Y	Y	Y	Y	N/A

Table 4 Statistics by industry

Figure 4 shows the RR distributions by industry. Although the shapes are similar, there are few differences as shown in Table 4. First, 10-years CDS for Finance industry has a much-concentrated distribution around 30% recovery rate compared to other industries. Also, its standard deviation is the lowest in both 5 and 10-year CDS and the second lowest in 1-year CDS. Second, 1-year CDS for Services has the smallest skewness despite having the highest standard deviation.

To further verify whether industry has different mean, *t*-test has been used for testing difference in means between each industry and whole sample without that industry. Null hypothesis is the above two series have the same mean. The test is conducted without assuming the series have equal variances. The results are reported in "*p*-value" and "Reject?" rows. For 5 and 10-year CDS, all industries have rejected the null hypothesis with less than 1% statistically significance. However, for 1-year CDS, we cannot reject the null hypothesis in 5% significant level except the services industry. Shorter maturity RR maturities may be more responsive to common economic factor and market sentiment instead of the industry specific factor.

In time series

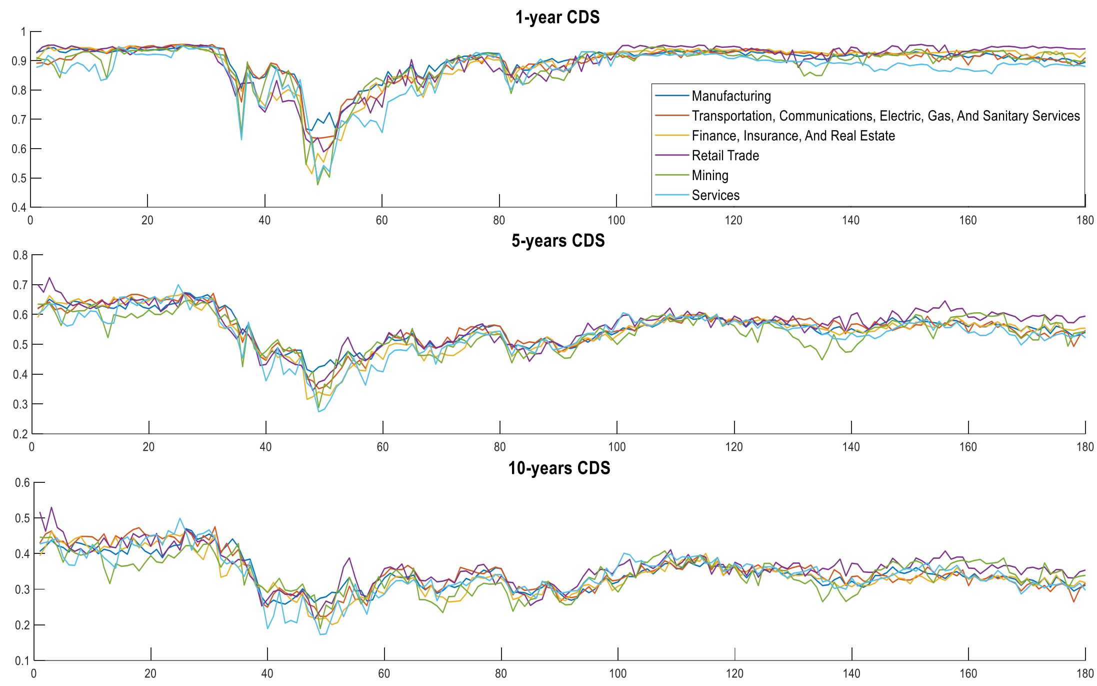


Figure 5 Recovery Rate by industry

Figure 5 shows the times series of recovery rates by industry. During the financial crisis 2008-2009, all industries experienced distress and their recovery rates fall together. However, the magnitudes are quite different. In 1-year CDS, Manufacturing enjoyed around 20% implied recovery rate more than mining industry, which stayed at the 50% level during the crisis. There is a drop only for Mining around 2015-2016 while other series remain stable. It is due to a slump in commodity markets during that period. To summarize, recovery rates demonstrate their dynamic feature in various business cycles. Besides, recovery rates among industries behave differently especially during recession/industry distress.

2.3 Monthly Beta estimation

The beta distribution is a good candidate for modelling recovery rates. The implied recovery rate is bounded between 0 and 1 although historical recovery rates could reach above 1 in few exceptions. The pdf, $f(x{:}\alpha,\beta)$ for this two-parameter distribution on [0, 1] is given by:

$$f(x{:}\,\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,dx}$$

The parameters α and β could be estimated by matching the mean (m) and variance (s) of the sample recovery rate, respectively:

$$\alpha = \frac{m^2(1-m) - ms}{s}$$

$$\beta = \frac{(1-m)(m^2 - m + s)}{s}$$

One of the reasons why the beta distribution is used is that it can create various distribution shapes by altering the two parameters, α and β.

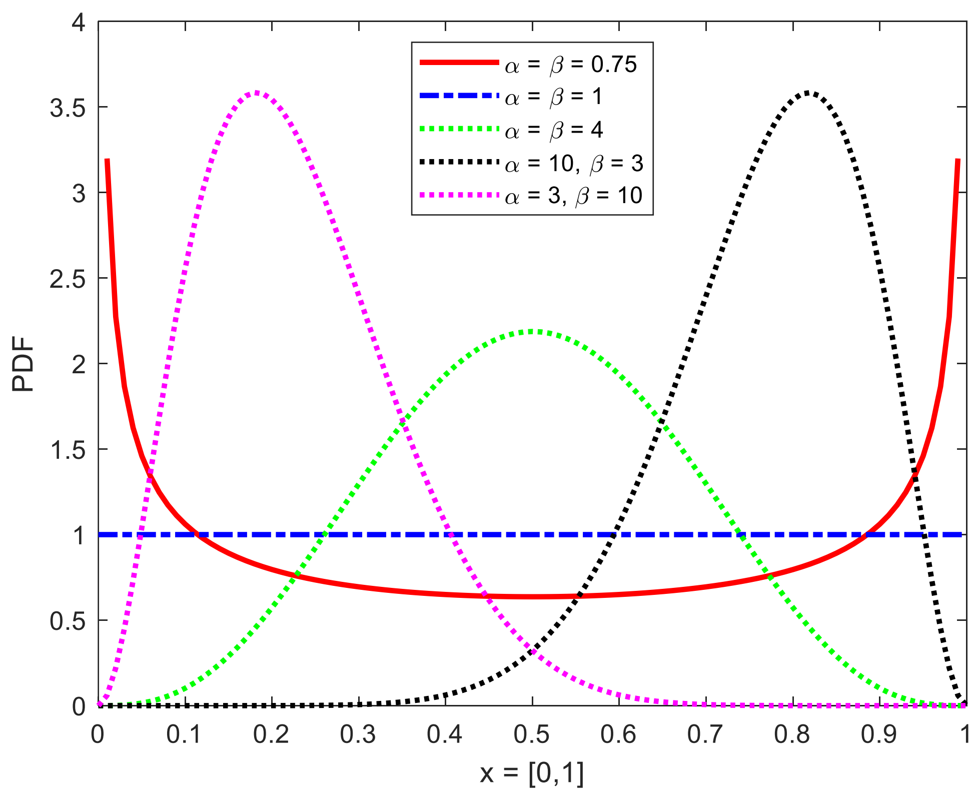


Figure 6 Beta distribution with different α and β

To enhance the accuracy of the fit with the beta distribution, I minimize the Root Mean Square Error (RMSE).

Steps are taken as follows for each cross section:

1. Divide the data into histograms with number of bins = 50, i.e., the width of bins is 0.02.
2. Compute the frequency within each bin as count of bin / number of total observations.
3. Calculate the observed PDF for each bin, frequency / bin's width.
4. Use the bin's center, i.e., for [0,0.02] bins, use 0.01 to calculate the theoretical PDF.
5. Solve the following optimization program for RMSE, $f_j^{hat}(x\colon \alpha, \beta)$ is the calculated pdf for each bin, $f_j^{thoe}(x\colon \alpha, \beta)$ is the theoretical beta pdf at bin center j.

$$\min_{\alpha,\beta} RSME = \min_{\alpha,\beta} \sqrt{\frac{\sum_{j=1}^{50}(f_j^{hat}(x\colon \alpha,\beta) - f_j^{thoe}(x\colon \alpha,\beta))^2}{50}}$$

Initial guess to α and β for optimization using moment matching are used to improve convergence and computation speed.

Step (1) – (5) are repeated until the end of time series. In total, 180 months of α and β are estimated.

Descriptive statistics for monthly RSME estimation

Here are the statistics for RSME fitting. The bracket next to industry is the number of firms in samples.

	Manufacturing (134)	Transportation (64)	Finance (57)	Retail Trade (28)	Mining (25)	Services (24)	Whole sample (351)
1-year							
Max	2.38	2.53	2.08	2.74	2.96	2.94	2.33
Min	0.23	0.39	0.27	0.69	0.71	0.65	0.28
Mean	1.34	1.33	1.32	1.55	1.67	1.69	0.99
Std Dev.	0.54	0.49	0.31	0.37	0.41	0.48	0.55
5-year							
Max	1.10	1.47	1.61	1.99	2.27	2.44	0.88
Min	0.48	0.80	0.77	1.09	1.15	1.11	0.36
Mean	0.81	1.06	1.10	1.49	1.59	1.59	0.58
Std Dev.	0.12	0.13	0.15	0.17	0.20	0.21	0.10
10-year							
Max	1.11	1.56	1.75	2.08	2.17	2.32	0.98
Min	0.60	0.84	0.85	1.14	1.14	1.17	0.43
Mean	0.86	1.12	1.12	1.54	1.62	1.64	0.68
Std Dev.	0.09	0.14	0.12	0.19	0.19	0.21	0.09

Table 5 Statistics for RSME

Table 5 reports the statistics for RSME fitting. Two findings can be observed. First, the RSME is monotonically decreasing with sample size since the continuous distribution can more closely match the distribution of population. Second, the RSME is decreasing with the maturity. As Figure 4 showed, longer maturity RR has a flatter distribution. Any outlier would have a smaller impact on the RSME. Overall, the RSME for pdf is acceptable as its average is under 2. Below figures show fitting of beta distribution to the whole samples, best and worst fit months are reported.

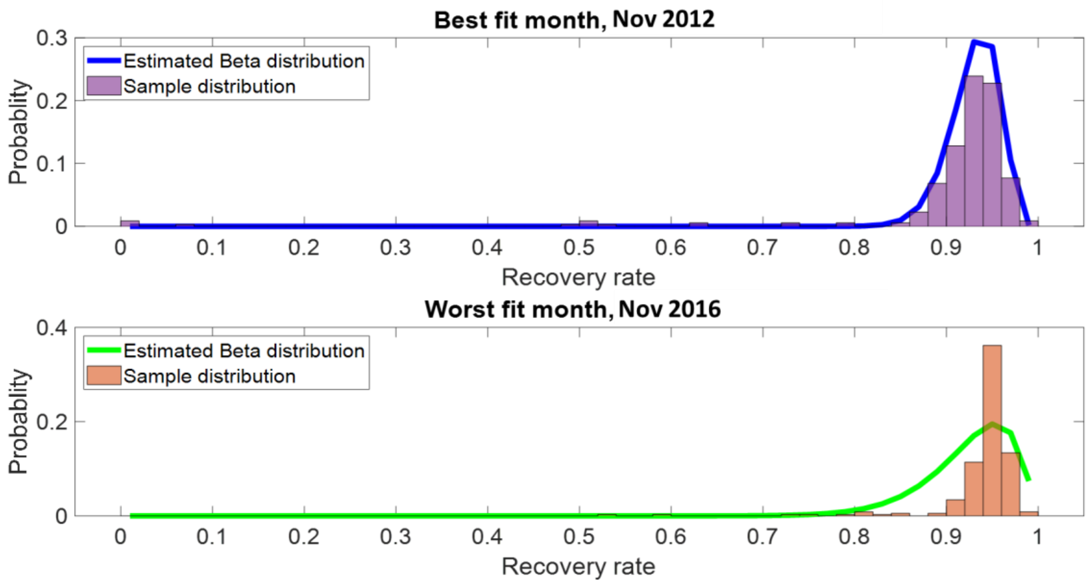


Figure 7 1-year CDS for Whole sample

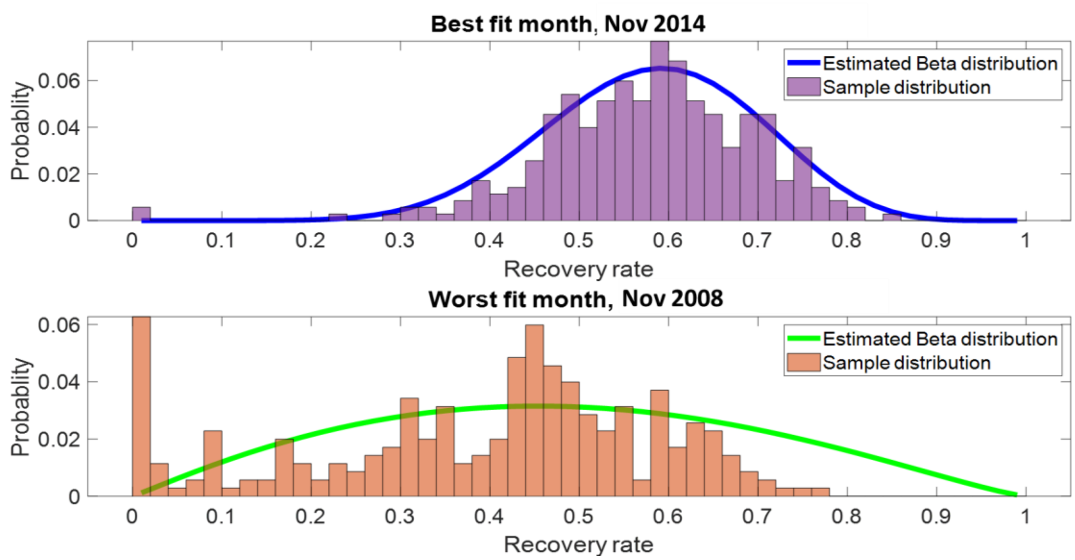


Figure 8 5-year CDS for Whole sample

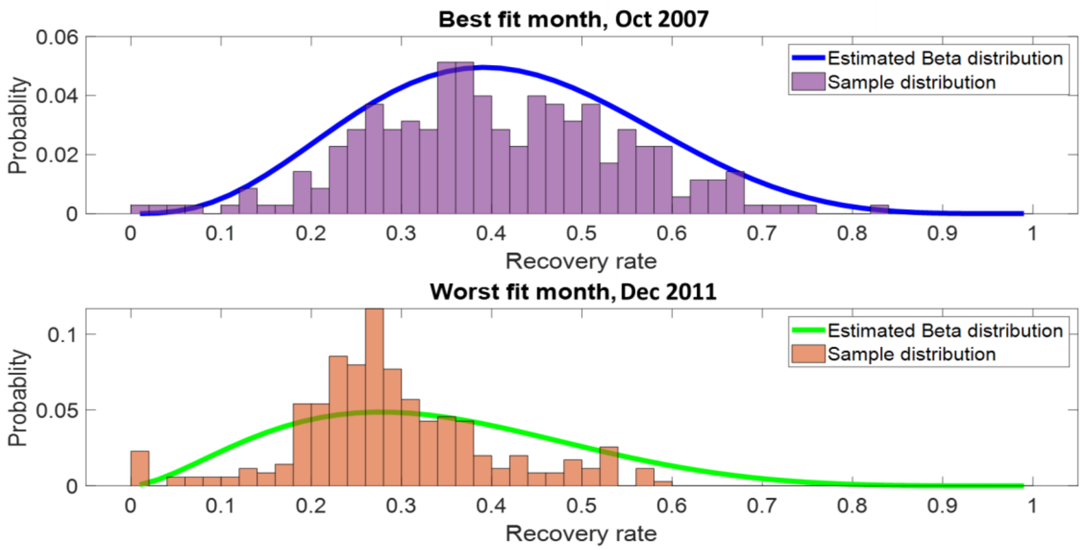


Figure 9 10-year CDS for Whole sample

Descriptive statistics for mean and precision

To make the model easier to interpret, α and β are transformed into their parameterized version in terms of the mean μ and the precision parameter ω as in Ferrari and Cribari-Neto (2004).

$$\mu = \frac{\alpha}{\alpha + \beta}$$

$$\omega = \alpha + \beta$$

The figures below show the effect of different μ and ω to the shape of the beta distribution.

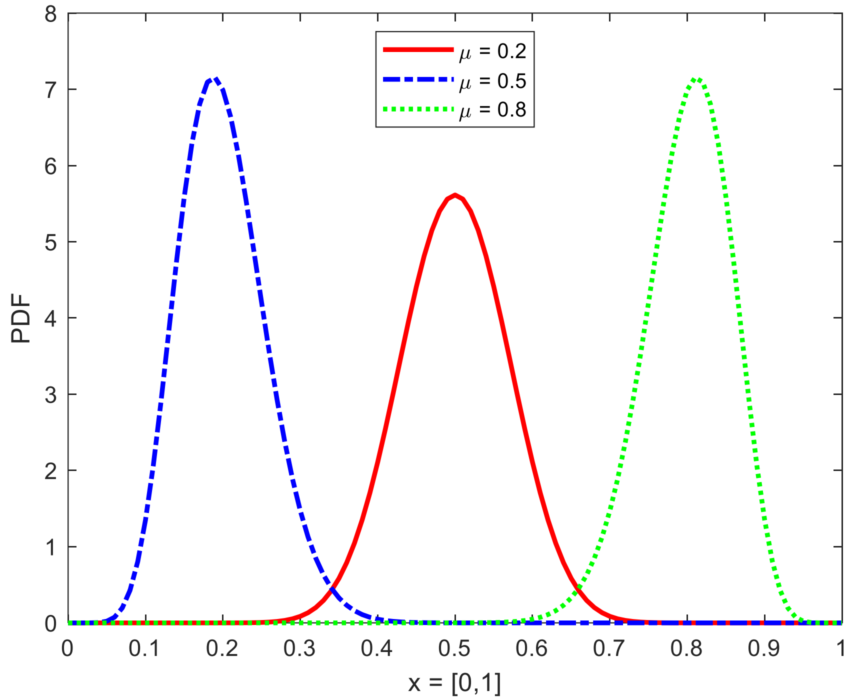


Figure 10 Beta distributions with ω = 50

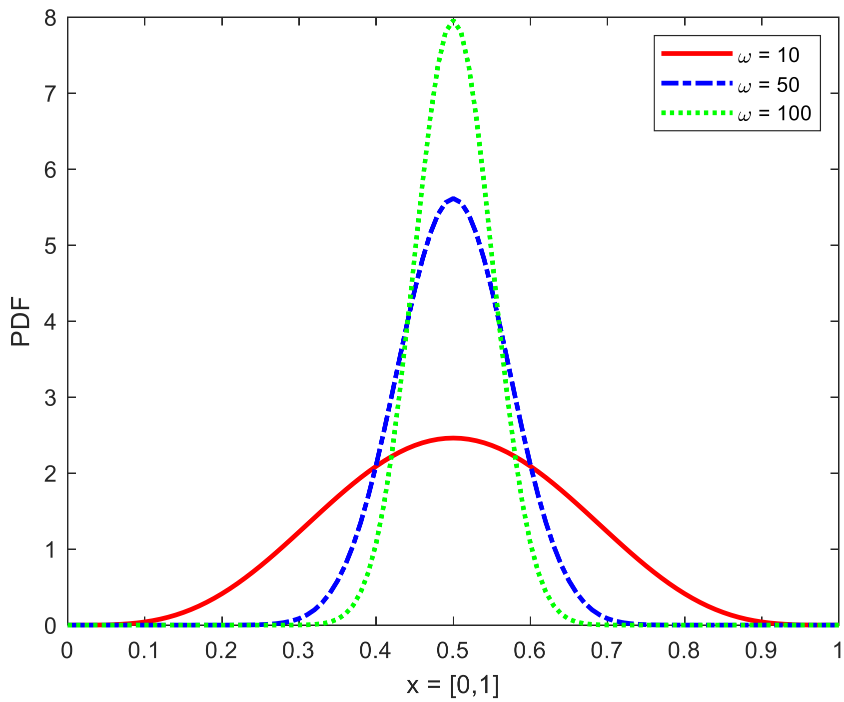


Figure 11 Beta distributions with μ = 0.5

	Manufacturing (134)	Transportation (64)	Finance (57)	Retail Trade (28)	Mining (25)	Services (24)	Whole sample (351)
1-year horizon							
Max	0.97	0.98	0.99	0.99	0.99	0.99	0.95
Min	0.41	0.37	0.38	0.35	0.38	0.00	0.46
Mean	0.86	0.87	0.83	0.82	0.89	0.89	0.90
Std Dev.	0.14	0.15	0.18	0.21	0.15	0.16	0.09
5-year horizon							
Max	0.71	0.73	0.74	0.81	0.80	0.85	0.67
Min	0.42	0.39	0.36	0.38	0.34	0.33	0.45
Mean	0.55	0.57	0.53	0.60	0.59	0.58	0.56
Std Dev.	0.06	0.07	0.09	0.10	0.11	0.10	0.05
10-year horizon							
Max	0.51	0.57	0.59	0.65	0.63	0.68	0.52
Min	0.22	0.19	0.23	0.20	0.13	0.16	0.25
Mean	0.39	0.44	0.40	0.45	0.43	0.40	0.38
Std Dev.	0.06	0.07	0.08	0.10	0.12	0.12	0.05

Table 6 Statistics for μ

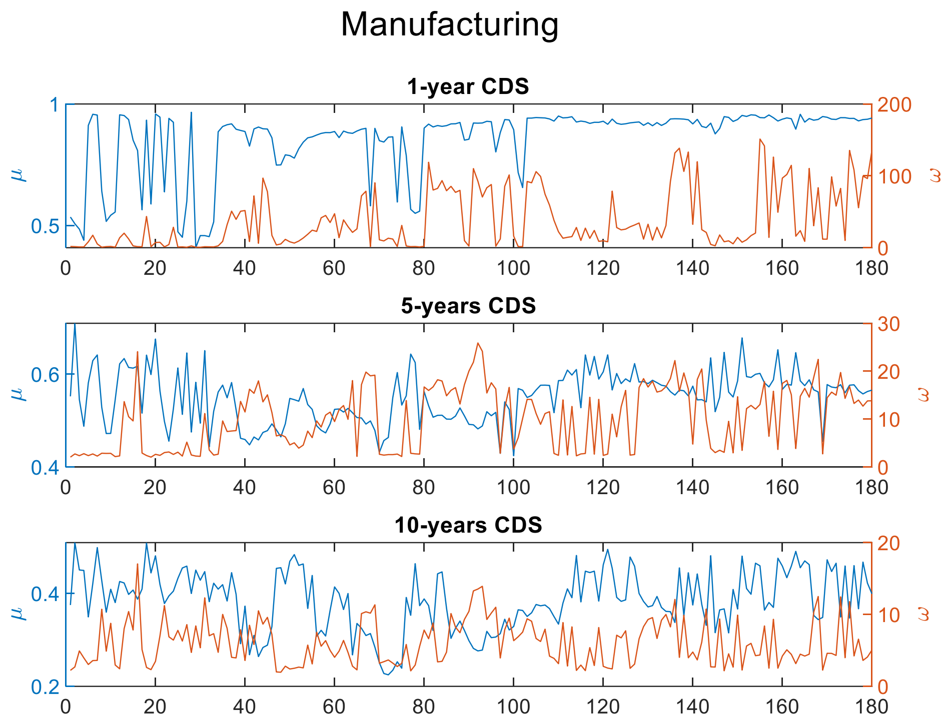
	Manufacturing (134)	Transportation (64)	Finance (57)	Retail Trade (28)	Mining (25)	Services (24)	Whole sample (351)
1-year horizon							
Max	151.21	275.97	155.12	270.25	1172.64	951.74	154.98
Min	0.62	0.45	0.53	0.25	0.47	0.24	1.02
Mean	38.15	46.11	19.32	30.69	32.21	45.62	74.36
Std Dev.	40.48	54.09	27.49	49.01	92.44	81.47	39.83
5-year horizon							
Max	25.89	32.49	29.49	29.10	161.71	31.50	27.25
Min	2.07	0.89	0.84	0.74	1.09	0.49	1.56
Mean	10.17	9.22	6.36	6.16	7.03	6.48	14.45
Std Dev.	6.48	7.06	6.45	5.10	13.01	6.58	5.64
10-year horizon							
Max	17.04	11.11	19.77	9.07	34.71	22.87	16.72
Min	1.97	0.62	0.84	0.49	1.16	0.75	1.26
Mean	6.13	4.17	3.25	2.98	4.14	3.82	8.65
Std Dev.	3.12	2.04	2.33	1.28	3.45	3.12	2.92

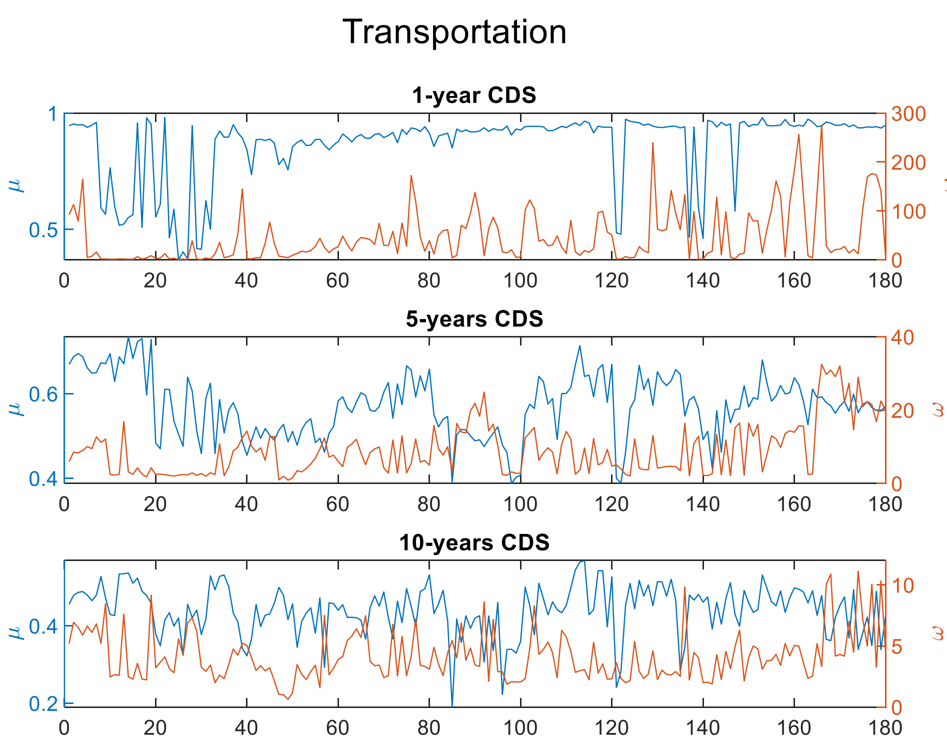
Table 7 Statistics for ω

Table 6 and Table 7 report the statistics for μ and ω, the estimated parameters of beta distribution are in line with Figure 4. The mean, μ, is decreasing with maturity and the difference among industry over the whole period is small. For precision, ω, it also shows a declining trend with maturity. As the shape of distribution is flatter, the precision drops. Industry-wise speaking, mining industry has the highest precision, this could be due to the common exposure of commodity price or just due to the small sample issue.

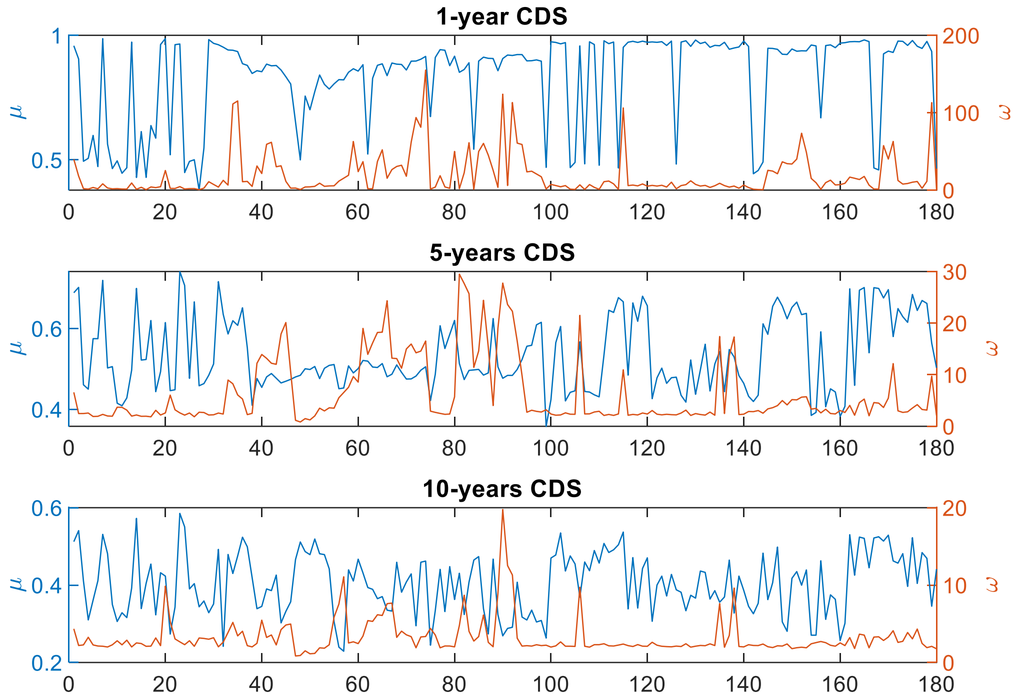
Appendix A1 shows the statistics for μ and ω using moment matching. For 5-year and 10-year CDS, their performances are similar comparing to RSME optimization. However, this method suffers a lot for 1-year CDS, ω have a std. dev. larger than to its mean, 134%. In Table 7 ω using RSME optimization, its 1-year result is much better, just 50% for std. dev. to mean ratio. Therefore, RSME optimization is better than moment-matching method and is adapted to proceed for time series.

In time series

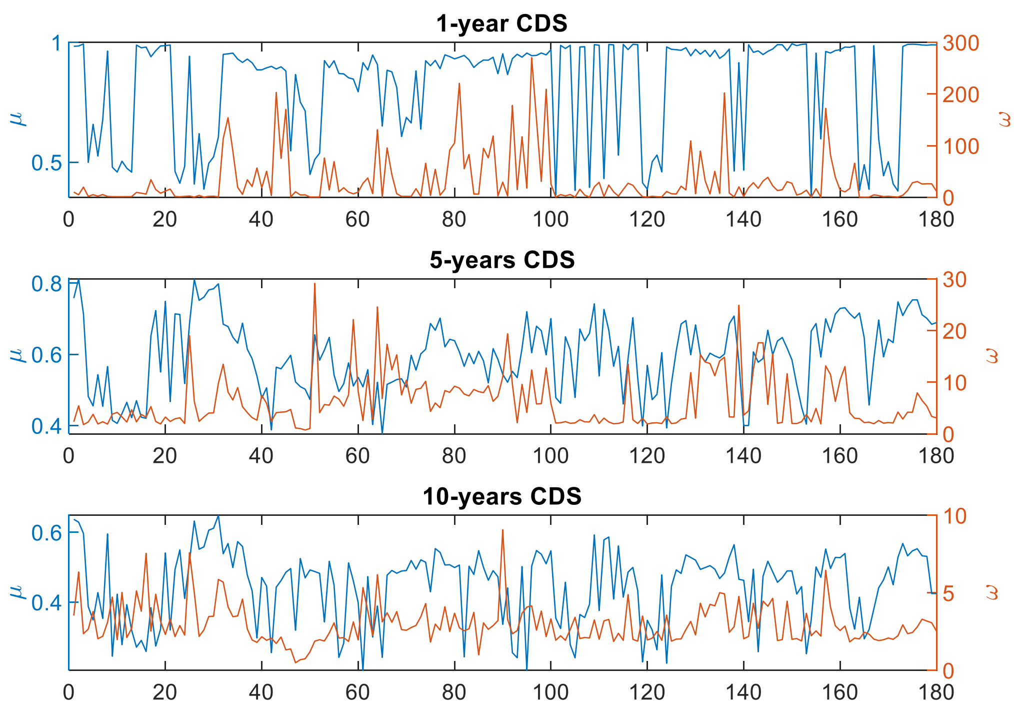




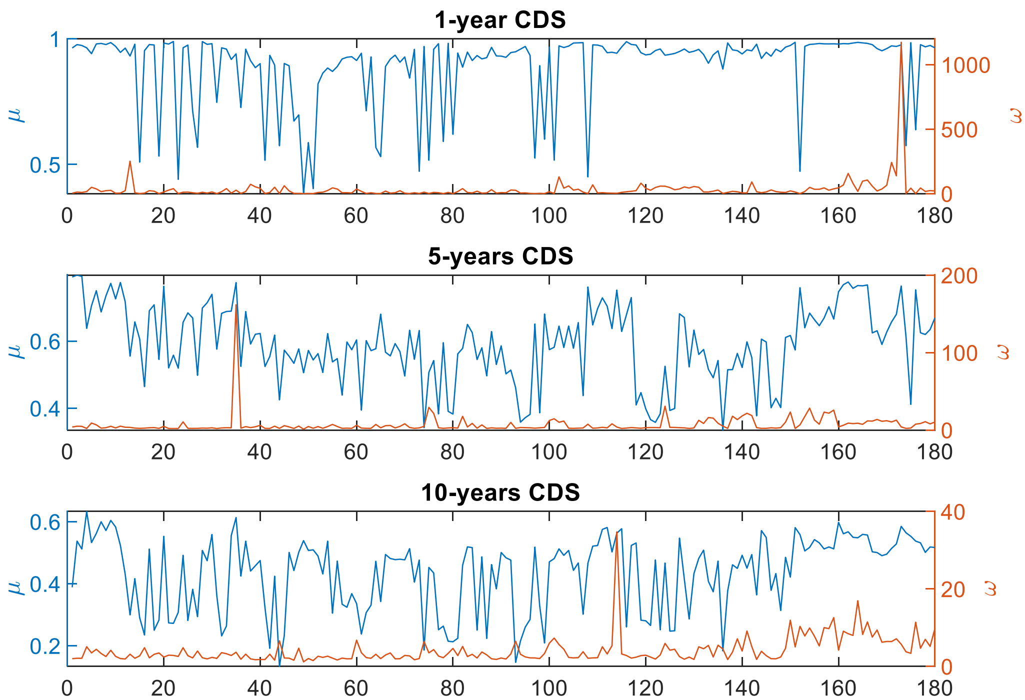
Finance



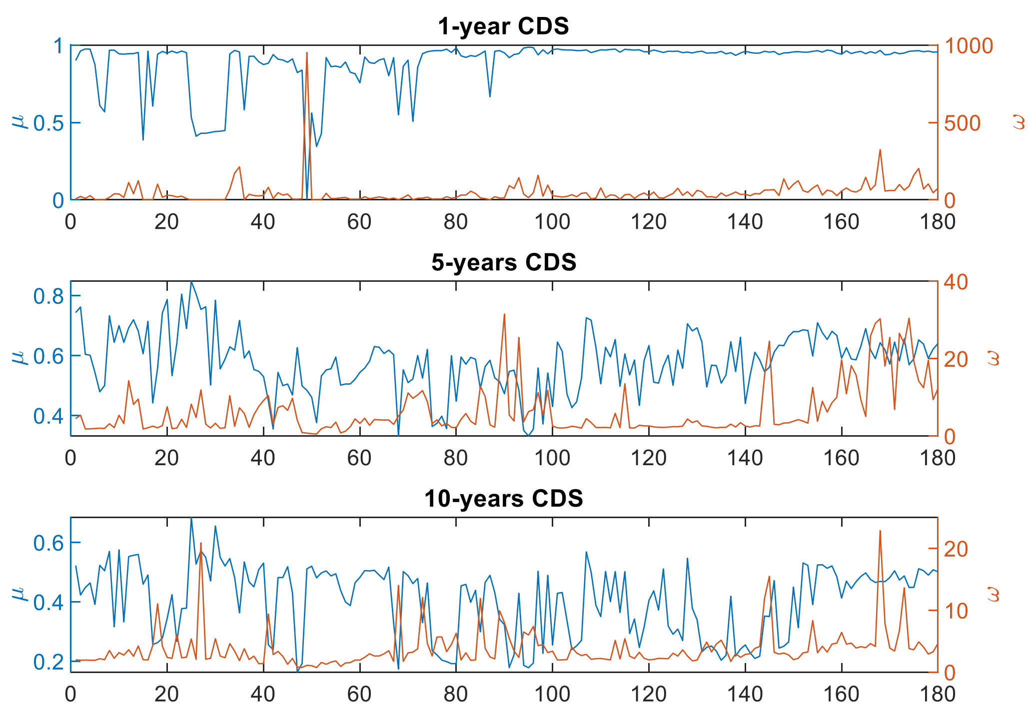
Retail Trade



Mining



Services



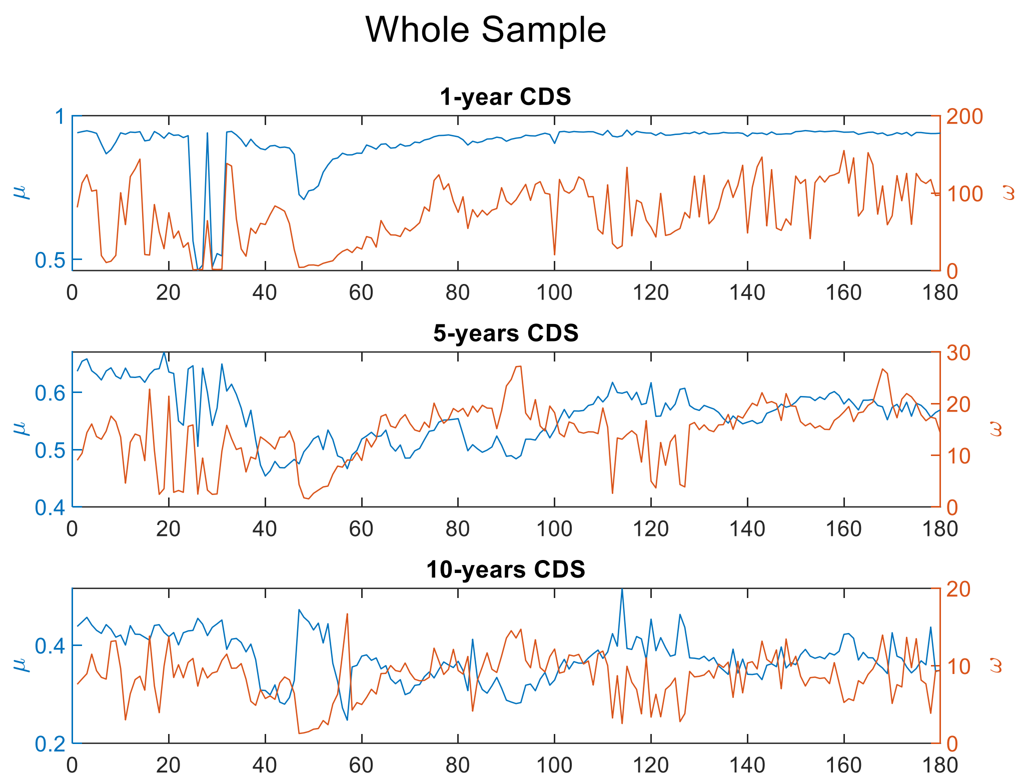


Figure 12 Mean and Precision in time series for whole sample and by industry

Figure 12 shows how mean μ and precision ω evolve over the sample period. Comparing to Figure 5, the industry mean looks more volatile in graph. It is not true however, with Table 4 and Table 6 , comparing the Std. Dev. to mean ratio, it is in fact lower for estimated mean. For whole sample, ranging from 8% - 13% vs 16% - 39% in various maturities. To further investigate the behavior of these parameters, a time series analysis will be conducted in next section.

Chapter 3
Time-series analysis of beta distribution coefficients

To begin with, I conduct the Augmented Dickey-Fuller test for a unit root to the series. After that, cointegration test will be applied. After applying the suitable differencing if any, VAR model is used for parameter mean μ and the precision parameter ω. Finally, in-sample fitting takes place.

3.1 Unit root test

The test for unit root test is the Augmented Dickey-Fuller (ADF) test. As the time series in previous section has no trend behavior and non-zero mean, the corresponding model variant[3] is adapted. 95 % Confidence level is applied in whether rejecting null hypothesis or not. The order of integration test is starting from I(3) to lower order, i.e., we test I(3) against I(2), if we reject null hypothesis, we test I(2) against I(1): If we cannot reject, we use I(2) as we cannot reject the null hypothesis. The sample size is 180, Schwert (1989) recommends a maximum lag $l_{max} = \left[12\left(\frac{T}{100}\right)^{0.25}\right]$,where T is the sample size. Therefore, I use $l_{max} = \left[12\left(\frac{180}{100}\right)^{0.25}\right] \approx 14$. Therefore, lag 0 and up to lag 14 are tested. Afterwards, the selection for p for t-statistics is based on BIC.

[3] If the differencing of time series has zero mean, zero mean variant is applied.

Industry	**Manufacturing**		**Transportation**		**Finance**		**Retail Trade**		**Mining**		**Services**		**Whole**	
Parameter	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
I(3) vs I(2)														
Lag	6	14	5	14	11	14	6	14	7	14	13	14	3	14
***p*-value**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Reject?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I(2) vs I(1)														
Lag	4	14	1	14	1	14	2	14	2	14	0	14	3	14
***p*-value**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Reject?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I(1) vs I(0)														
Lag	0	14	2	14	0	14	0	14	1	14	1	14	4	14
***p*-value**	0.001	0.232	0.006	0.310	0.001	0.027	0.001	0.056	0.001	0.985	0.001	0.417	0.004	0.380
Reject?	Y	N	Y	N	Y	Y	Y	N	Y	N	Y	N	Y	N

Table 8 1-year CDS ADF test

Industry	**Manufacturing**		**Transportation**		**Finance**		**Retail Trade**		**Mining**		**Services**		**Whole**	
Parameter	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
I(3) vs I(2)														
Lag	5	14	4	14	5	14	4	14	4	14	6	14	2	14
***p*-value**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Reject?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I(2) vs I(1)														
Lag	2	14	0	14	1	14	3	13	1	14	0	14	1	14
***p*-value**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Reject?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I(1) vs I(0)														
Lag	1	14	1	14	0	14	0	14	2	14	1	14	2	14
***p*-value**	0.001	0.045	0.001	0.877	0.001	0.220	0.001	0.120	0.003	0.020	0.001	0.599	0.145	0.177
Reject?	Y	Y	Y	N	Y	N	Y	N	Y	Y	Y	N	N	N

Table 9 5-year CDS ADF test

Industry	Manufacturing		Transportation		Finance		Retail Trade		Mining		Services		Whole	
Parameter	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
I(3) vs I(2)														
Lag	4	14	5	14	5	14	5	14	5	14	7	14	4	14
***p*-value**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Reject?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I(2) vs I(1)														
Lag	0	14	0	13	3	6	1	14	1	10	0	7	0	14
***p*-value**	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Reject?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
I(1) vs I(0)														
Lag	1	13	0	14	0	0	1	2	1	11	1	8	0	10
***p*-value**	0.001	0.020	0.001	0.505	0.001	0.001	0.001	0.001	0.001	0.506	0.001	0.051	0.001	0.164
Reject?	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	N	Y	N

Table 10 10-year CDS ADF test

Tables 8 to 10 show the ADF test results for different industries and maturities. In each time series across industries, the lag selected are different. The selection of lag of each series is based on BIC, therefore, each parameter, maturities and industries could have different lag for ADF test. Almost all the μ do not show any existence of unit root, this is expected as the mean should be stationary in a long run. On the other hand, some ω show that there is a unit root and should apply differencing in level 1. However, we do not expect the change of the precision to be stationary.

3.2 Cointegration

If any μ and ω pair both do not have I(d) with d >1, there is no cointegration vector that could reduce the order of integration. Among 21 pairs, there is only one pair with I(1). No cointegration test is conducted as VAR in level will be used in this paper.

3.3 VAR estimation

In this paper, mean and precision parameters time series are being modelled. Financially speaking, both mean and precision of implied recovery rate, like asset return, are expected to be an order of integration of 0. Although the Augmented Dickey-Fuller tests in 3.1 do show some of the series have a unit root, there is a risk of over differencing. In fact, Sims, Stock & Watson (1990) advised against differencing even if there is a unit root. Phillips (1998) reports that VAR in level estimation is a consistent estimator in impulse-response function in short and medium run. Therefore, the estimation is conducted under VAR in level.

For VAR estimation, maximum likelihood approach is used for estimation. The VAR(*p*), with *p* lags, has the general form:

$$\begin{pmatrix} \mu_t \\ \omega_t \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} + \begin{pmatrix} \Phi_{11}^1 & \Phi_{12}^1 \\ \Phi_{21}^1 & \Phi_{22}^1 \end{pmatrix} \begin{pmatrix} \mu_{t-1} \\ \omega_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \Phi_{11}^p & \Phi_{12}^p \\ \Phi_{21}^p & \Phi_{22}^p \end{pmatrix} \begin{pmatrix} \mu_{t-p} \\ \omega_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Where δ is constant, $\varepsilon \sim N(0,1)$. If the time series has unit root, $\begin{pmatrix} \mu_t \\ \omega_t \end{pmatrix}$ is replaced by $\begin{pmatrix} \Delta\mu_t \\ \Delta\omega_t \end{pmatrix}$

The selection of the optimal number of lags is made by checking their BIC and their residuals. To test whether residuals are white noise or not, the Ljung–Box(LB) test is applied (Ljung and Box, 1978) with the null hypothesis that residuals are independently distributed, i.e., white noise. The number of lags used in this test is also 14, which is the upper bound for testing. The lowest BIC is selected with at least 7 lags passing the Ljung-Box test in the μ and ω pair.

Industry	Manufacturing		Transportation		Finance		Retail Trade		Mining		Services		Whole	
Parameter	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
1-year CDS														
Lag select	1	1	5	5	1	1	2	2	2	2	2	2	6	6
No. of lag passing LB test	14	13	14	11	14	13	14	14	8	14	14	14	9	14
5-year CDS														
Lag select	2	2	2	2	3	3	3	3	2	2	4	4	3	3
No. of lag passing LB test	8	14	14	14	14	14	14	14	14	14	14	14	9	14
10-year CDS														
Lag select	2	2	2	2	1	1	2	2	1	1	2	2	2	2
No. of lag passing LB test	14	14	14	14	14	14	14	14	14	13	14	14	14	14

Table 11 VAR lag selection and LB test results

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	***t*-stat**	**Value**	***t*-stat**	**Value**	***t*-stat**
Constant(1)	0.336	5.137	0.042	2.051	0.025	0.958
Constant(2)	7.649	0.198	4.852	1.296	4.096	1.819
AR{1}(1,1)	0.631	7.890	0.578	8.335	0.699	6.982
AR{1}(2,1)	45.745	0.969	-2.918	-0.233	-2.796	-0.323
AR{1}(1,2)	0.000	1.604	0.000	0.034	0.003	2.273
AR{1}(2,2)	0.378	4.689	0.562	7.738	0.361	3.601
AR{2}(1,1)	0.061	0.655	-0.001	-0.016	0.169	1.695
AR{2}(2,1)	-36.116	-0.663	2.778	0.189	2.335	0.271
AR{2}(1,2)	0.000	1.499	-0.001	-1.525	0.000	0.108
AR{2}(2,2)	0.245	2.895	-0.019	-0.222	0.187	1.852
AR{3}(1,1)	-0.313	-3.790	0.340	4.959		
AR{3}(2,1)	-105.324	-2.165	-3.420	-0.276		
AR{3}(1,2)	0.000	0.138	0.001	2.306		
AR{3}(2,2)	0.062	0.716	0.261	3.593		
AR{4}(1,1)	0.660	7.981				
AR{4}(2,1)	160.453	3.289				
AR{4}(1,2)	0.000	-1.505				
AR{4}(2,2)	0.008	0.090				
AR{5}(1,1)	-0.180	-1.879				
AR{5}(2,1)	28.881	0.512				
AR{5}(1,2)	0.000	-0.179				
AR{5}(2,2)	-0.032	-0.382				
AR{6}(1,1)	-0.268	-3.330				
AR{6}(2,1)	-79.160	-1.664				
AR{6}(1,2)	0.000	1.877				
AR{6}(2,2)	0.064	0.798				

Table 12 VAR(p) Parameters for whole sample AR{p}(1,2) means the parameter for μ from p lags of ω

Table 12 shows the parameters for whole sample, Appendix A2 shows the industry-wise parameters. All parameters are positive except some of the parameter ω from μ. This is interesting that as this implies high mean could flatten out the distribution. Some parameters are not statistically significant. For forecasting purposes, the model will only use the 5 % statistically significant factors.

3.5 In-sample fitting

The in-sample fitting performance of the model is examined. First, I calculate the residuals between the empirical values and the VAR model values:

$$\varepsilon_t = y_t^{Empirical} - y_t^{Estimated}$$

After that, the RSME is computed as follows

$$RSME = \sqrt{\frac{\sum_{j=1}^{180-p} {\varepsilon_t}^2}{180-p}}$$

Industry	**Manufacturing**		**Transportation**		**Finance**		**Retail Trade**		**Mining**		**Services**		**Whole**	
Parameter	μ	ω	M	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
1-year CDS														
RSME	0.109	34.759	0.120	45.050	0.168	25.947	0.191	45.479	0.142	90.299	0.126	80.785	0.051	30.008
5-year CDS														
RSME	0.046	5.311	0.054	5.161	0.076	4.586	0.081	4.743	0.092	13.005	0.082	5.037	0.021	3.864
10-year CDS														
RSME	0.049	2.902	0.060	1.910	0.074	1.994	0.092	1.210	0.107	3.221	0.101	3.050	0.030	2.566

Table 13 RSME of in-sample fitting

When the number of sample firms in the industry decreases, the RSME increases. This is promising as if we could increase more data when estimating the beta distribution, we could have a better modelling result in the time series.

Some industries like transportation and mining have a higher RSME comparing to others, this may be due to some unknown industry factors which have not been included in our data.

Comparing the parameters to its average (Table 6 and Table 7), estimation for μ is much better than ω. For example, in whole data, the average for μ is 0.9 with RSME 0.051 while ω is 74 with RSME 30. Finally, 5-year CDS have a lower RSME comparing to 10-year ones. As the average of parameters 5-year CDS are higher than those in 10-year, 5-year CDS have the best estimation errors among the three maturities. These patterns could be observed graphically too. Below are graphs showing the time series of empirical vs estimated VAR model for the whole sample. For the remaining 6 industries, they are put under Appendix A3.

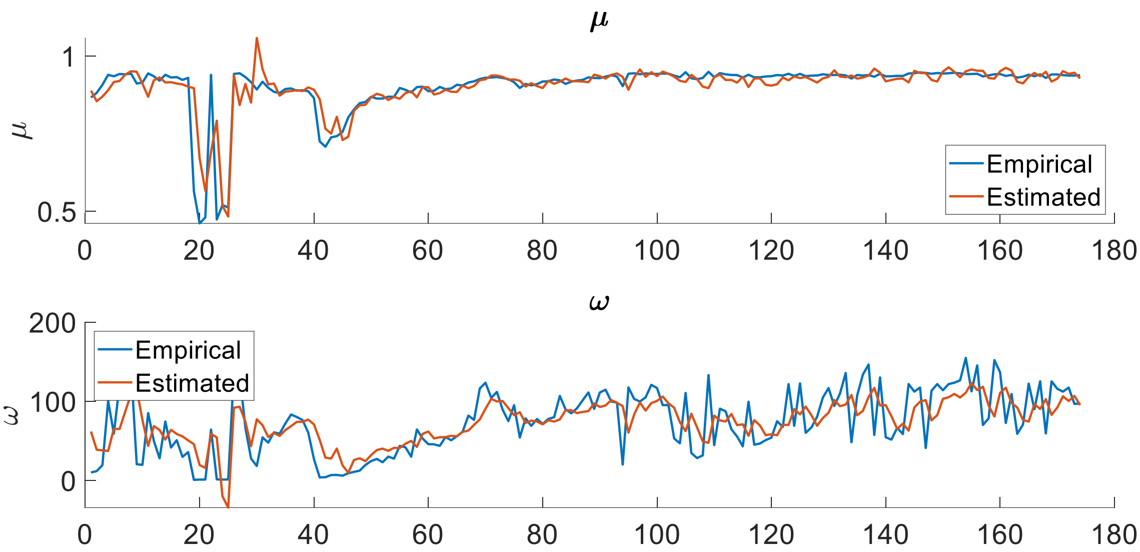


Figure 13 Empirical vs Estimated for 1-year CDS

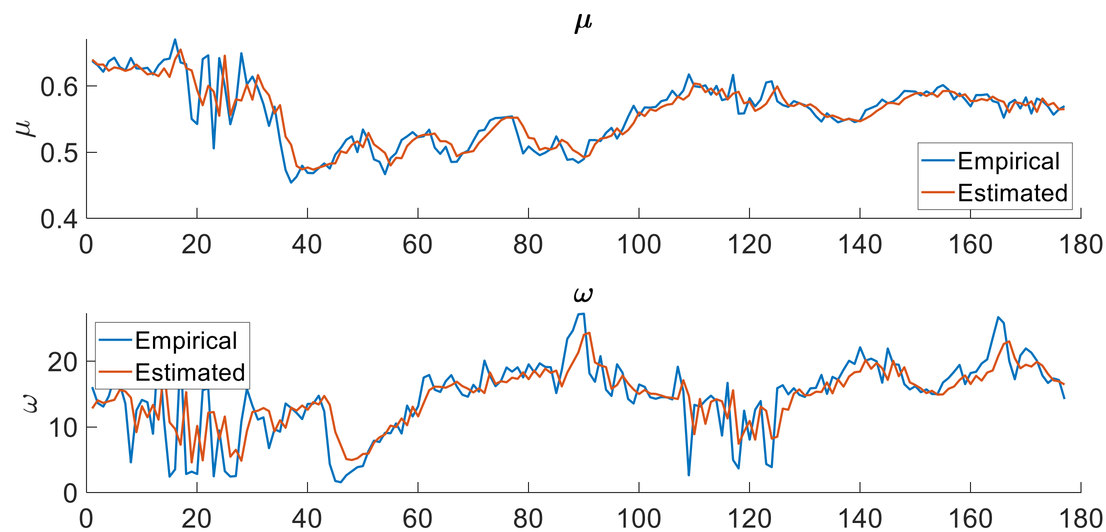


Figure 14 Empirical vs Estimated for 5-year CDS

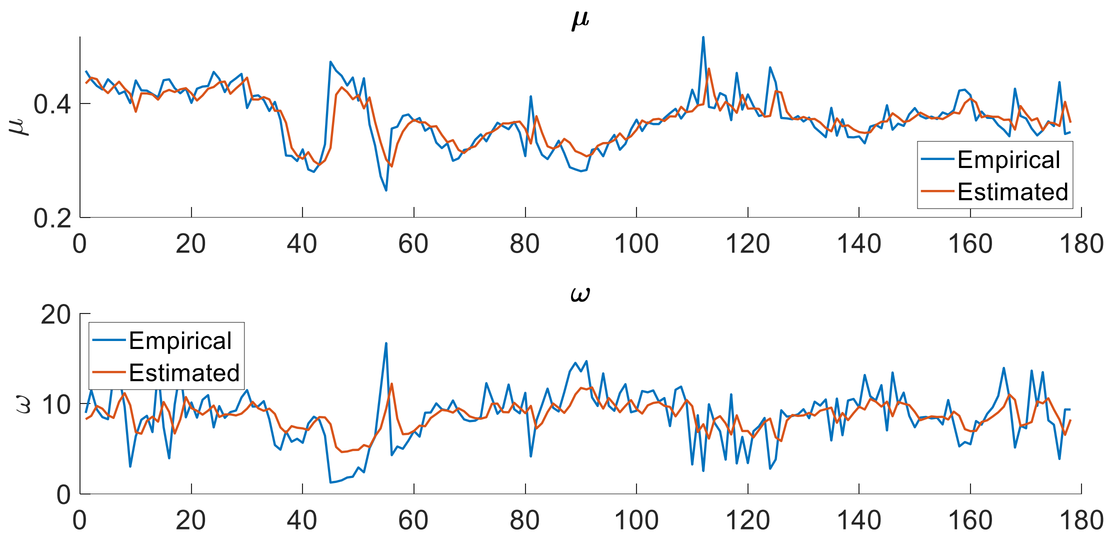


Figure 15 Empirical vs Estimated for 10-year CDS

Chapter 4
Out-of-sample tests

To further examine the quality of the models, pseudo out-of-sample forecast is conducted. Please note that only statistically significant factors are included in forecasting.

4.1 Procedure

1. Following steps are carried out. ,Starting with *n* number of samples to estimate the parameters in VAR(p). In our cases, 90 samples are used in first estimation. The data set denotes as I_n

2. Forecast 6 periods ahead. $Y_{n+h}^{Forecast} = E(Y_{n+h}\ |I_n)\ for\ h = 1,2, \dots ,6$

3. Compute the forecast error, $e_{n+h}^{Forecast} = Y_{n+h}^{\text{Empirical}} - Y_{n+h}^{Forecast}, h = 1,2, \dots ,6$

4. Instead of a rolling window for estimation. I increase by 1 the number of data points used for estimating parameters, i.e., I_{n+1}, and repeat Steps (1) – (3).

5. Repeat Steps (1) – (4) until the end of samples.

6. A series of $\{e_{n+h}^{Forecast}\}$ is obtained.

7. Calculate the RSME of forecasting error for each forecast period *h*.

$$RSME(Y_{n+h}^{Forecast}) = \sqrt{\frac{\sum_{n=90}^{180-90-h}(e_{n+h}^{Forecast})^2}{180-90-h}}$$

4.3 Results

	Manufacturing		**Transportation**		**Finance**		**Retail Trade**		**Mining**		**Services**		**Whole**	
h	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
1	0.06	42.6	0.12	59.3	0.18	22.7	0.23	53.9	0.13	199.1	0.02	71.7	0.02	35.7
2	0.08	43.9	0.13	65.2	0.20	21.6	0.23	53.6	0.13	204.9	0.02	74.3	0.03	35.4
3	0.09	45.7	0.13	66.7	0.20	21.6	0.24	54.4	0.14	240.9	0.03	74.3	0.03	37.1
4	0.09	46.4	0.14	67.8	0.20	21.5	0.24	54.6	0.14	256.5	0.03	73.5	0.03	37.3
5	0.10	46.9	0.14	68.5	0.20	24.6	0.24	55.2	0.14	289.0	0.04	73.5	0.04	38.2
6	0.10	47.9	0.13	68.8	0.20	24.7	0.24	55.2	0.14	322.3	0.04	74.0	0.05	39.0

Table 14 RSME of forecast for 1-year CDS

	Manufacturing		**Transportation**		**Finance**		**Retail Trade**		**Mining**		**Services**		**Whole**	
h	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
1	0.04	6.50	0.05	6.46	0.09	4.58	0.09	4.98	0.10	10.61	0.08	6.73	0.01	3.50
2	0.05	6.75	0.06	7.12	0.09	4.66	0.10	4.95	0.11	10.64	0.08	6.89	0.02	4.14
3	0.05	6.66	0.07	7.73	0.10	4.70	0.10	4.84	0.12	10.62	0.09	6.88	0.02	4.12
4	0.05	7.01	0.07	8.09	0.10	5.18	0.10	5.40	0.12	10.69	0.08	6.89	0.02	4.27
5	0.05	7.14	0.07	8.82	0.10	5.33	0.10	5.68	0.13	10.71	0.08	7.55	0.02	4.37
6	0.05	7.12	0.07	9.20	0.10	5.20	0.10	5.83	0.13	10.77	0.08	7.48	0.02	4.50

Table 15 RSME of forecast for 5-year CDS

	Manufacturing		**Transportation**		**Finance**		**Retail Trade**		**Mining**		**Services**		**Whole**	
h	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω	μ	ω
1	0.05	3.29	0.07	2.21	0.07	2.01	0.09	1.03	0.10	4.87	0.10	3.08	0.03	3.11
2	0.05	3.26	0.07	2.36	0.07	2.15	0.09	1.04	0.12	4.95	0.10	3.16	0.03	2.94
3	0.05	3.46	0.07	2.43	0.08	2.40	0.10	1.00	0.11	4.97	0.11	3.16	0.03	3.49
4	0.05	3.38	0.07	2.63	0.08	2.55	0.10	1.00	0.11	4.97	0.11	3.16	0.03	3.31
5	0.05	3.41	0.07	2.62	0.08	2.56	0.10	1.01	0.11	4.97	0.11	3.15	0.03	3.62
6	0.05	3.42	0.07	2.61	0.08	2.60	0.10	1.01	0.11	5.01	0.11	3.15	0.03	3.51

Table 16 RSME of forecast for 10-year CDS

Table 14 – Table 16 report the quality of forecast, *h* is the number of steps in the forecast. The results show a similar pattern with the in-sample fitting. Transport and mining do not perform as good as the other industries.

The 5-year CDS has the best forecasting result among other maturities. Also, the sample size in beta distribution estimation does play a role in the quality of forecasting. The more firms included in the sample, the better forecasting results are. Surprisingly, the forecast quality for 5-year and 10-year CDS does not deteriorate much when forecasting horizon increases from 1 to 6, which is 6 months later.

To illustrate the forecasting quality, Figure 16 – Figure 21, are the empirical vs forecast parameters with different forecast horizon, h = 1,3 and 6, for whole samples. For 1-year CDS, it underestimates the mean of recovery rate, but this pattern does not show up in 5, 10-year CDS.

Overall, the forecast quality is good especially for the mean μ. The whole sample forecast error for 10-year CDS is 0.03 with its mean is 0.38, which is 8% of error. For 1-year and 5-year CDS, the maximum errors are 0.05 and 0.02. Their mean are 0.90 and 0.56, which implied smaller percentage, 5.6% and 3.6% respectively.

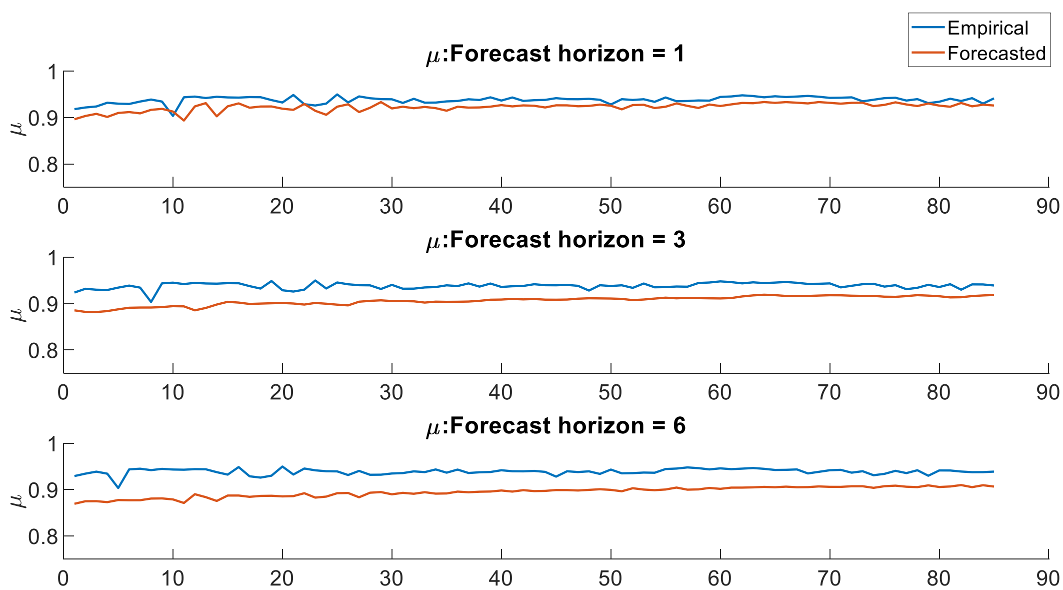


Figure 16 Forecast μ for 1-year CDS

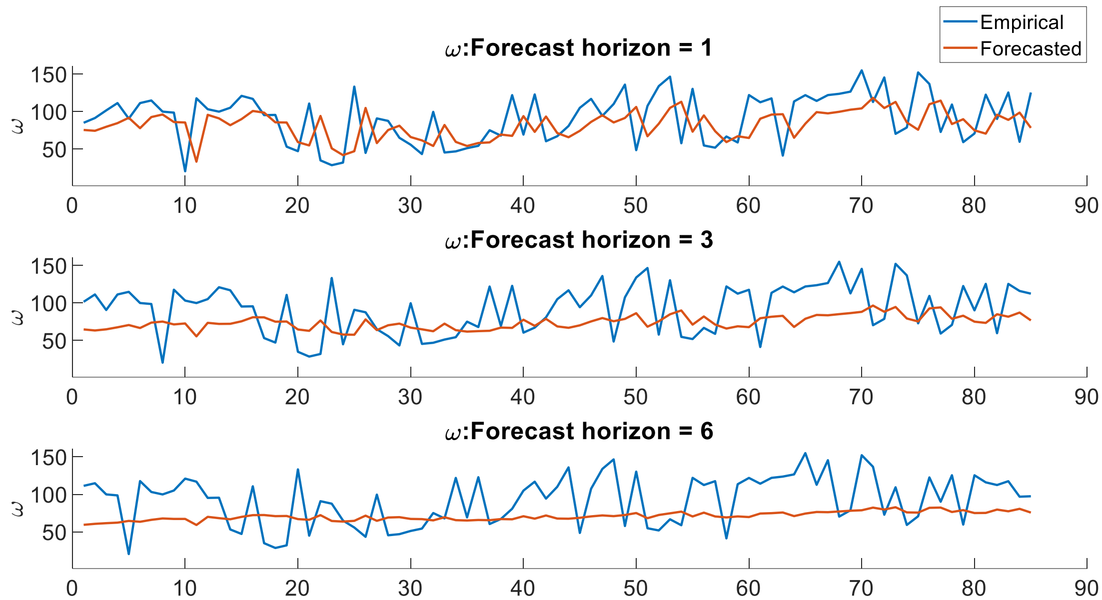


Figure 17 Forecast ω for 1-year CDS

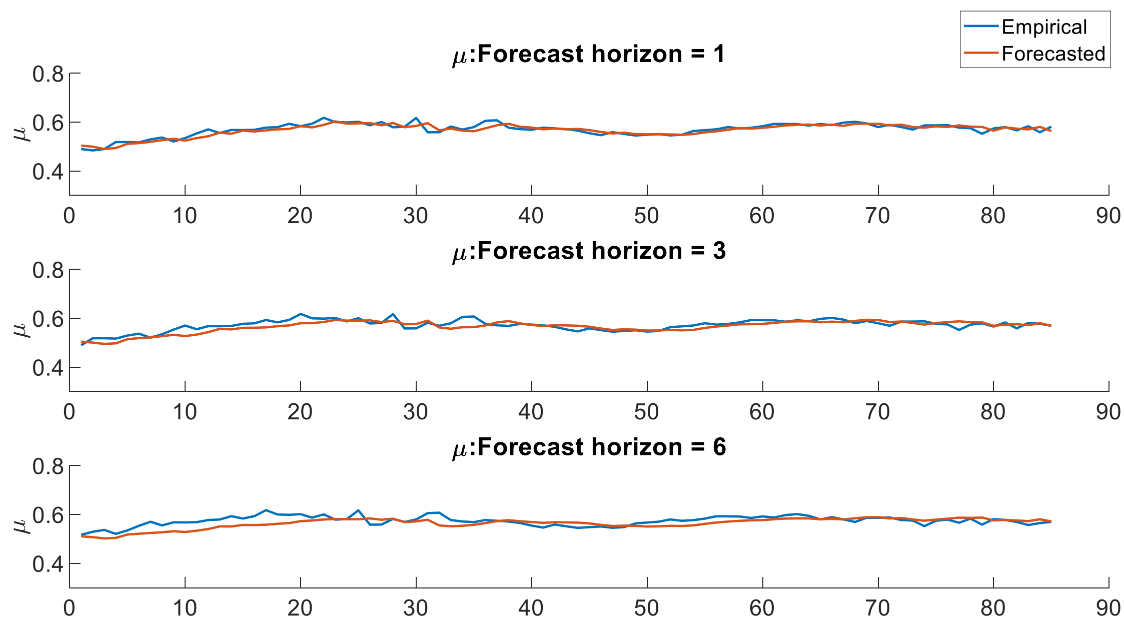


Figure 18 Forecast μ for 5-year CDS

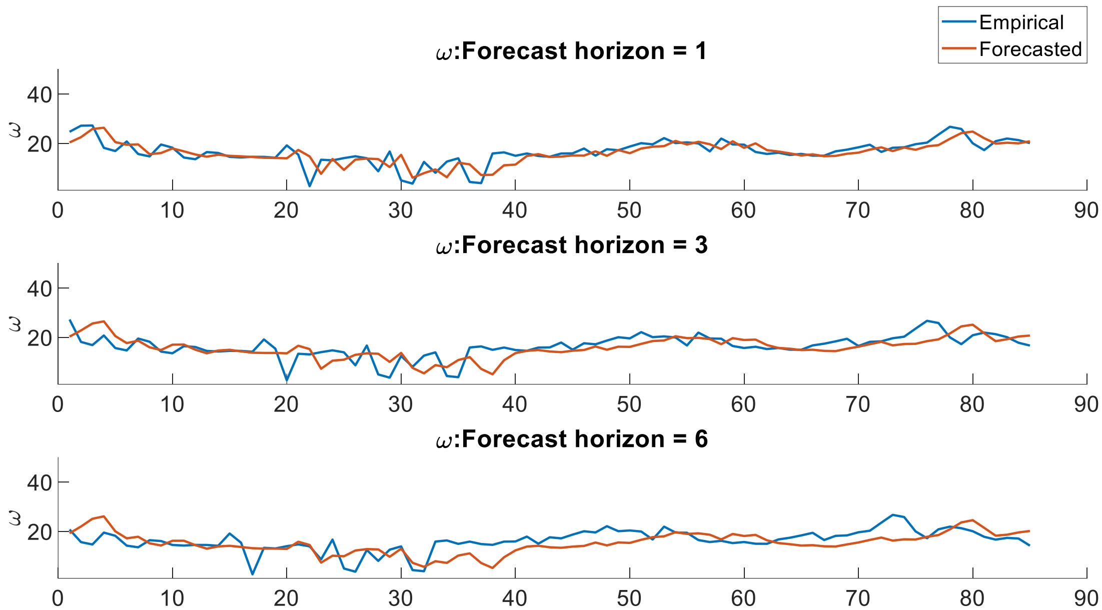


Figure 19 Forecast ω for 5-year CDS

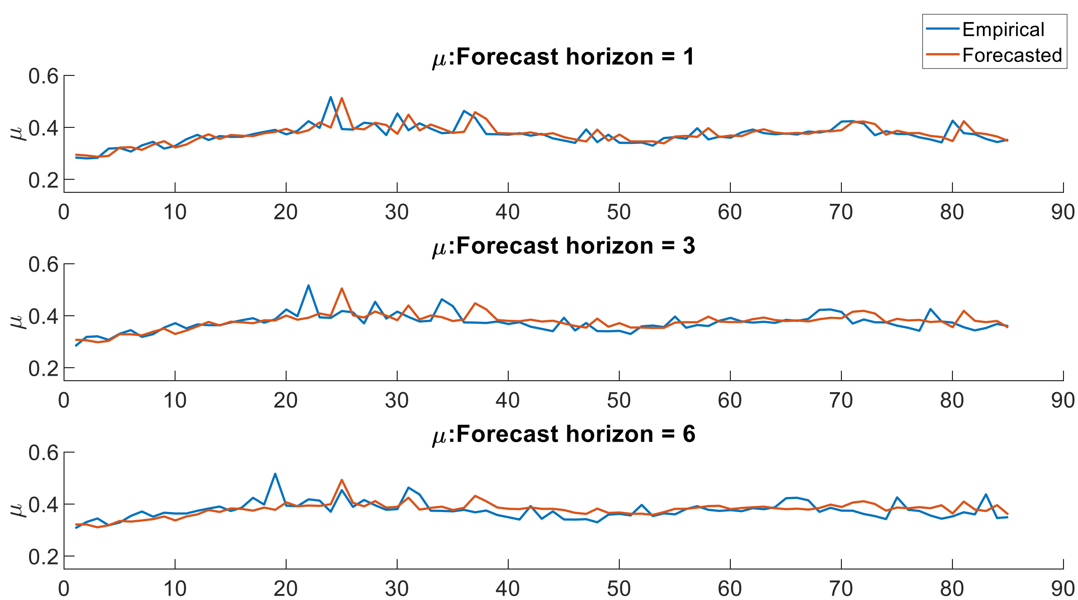


Figure 20 Forecast μ for 10-year CDS

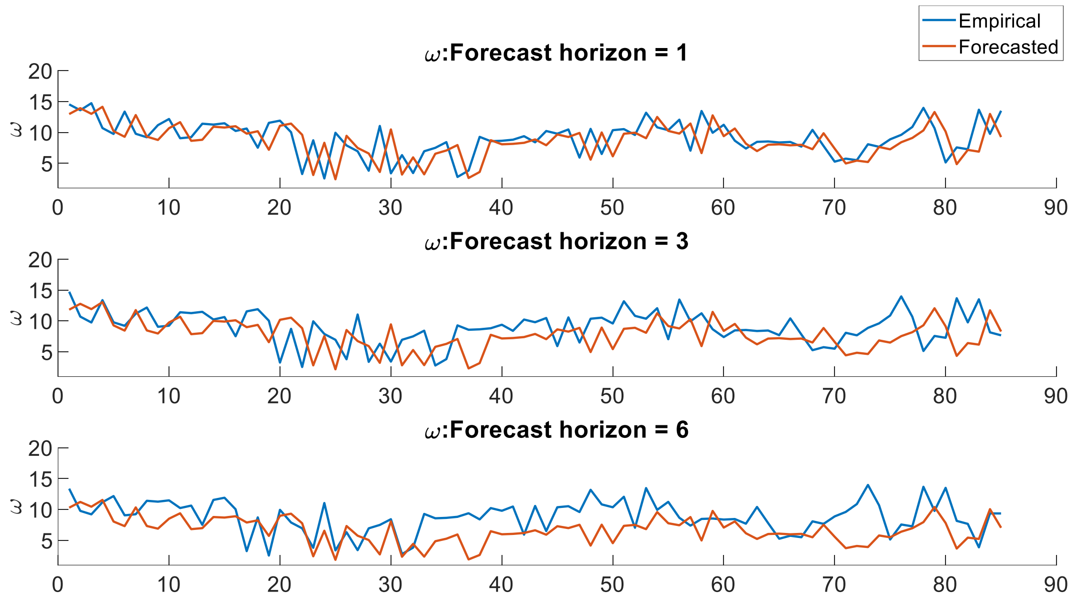


Figure 21 Forecast ω for 10-year CDS

To further examine the quality of forecasting the beta distribution, I plotted empirical vs forecast beta distribution with different forecast horizon, h = 1,3 and 6, for whole samples. Beta distributions are obtained by the average of empirical/forecast parameters in the forecasting period. All forecasts have a fatter tail than the empirical one, especially in 1-year CDS.

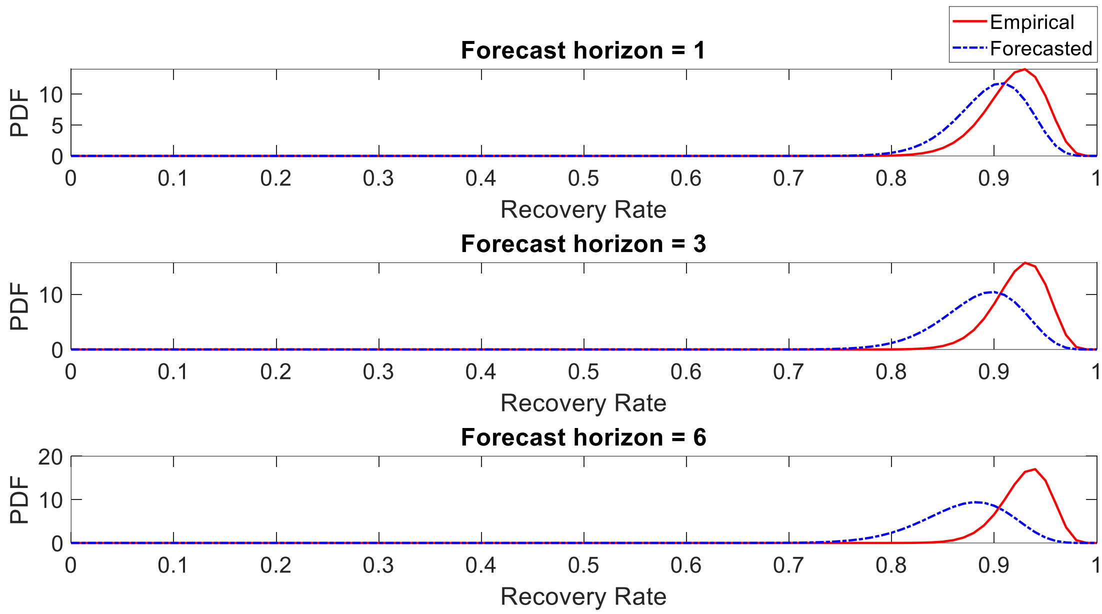


Figure 22 Empirical vs Forecast for 1-year CDS

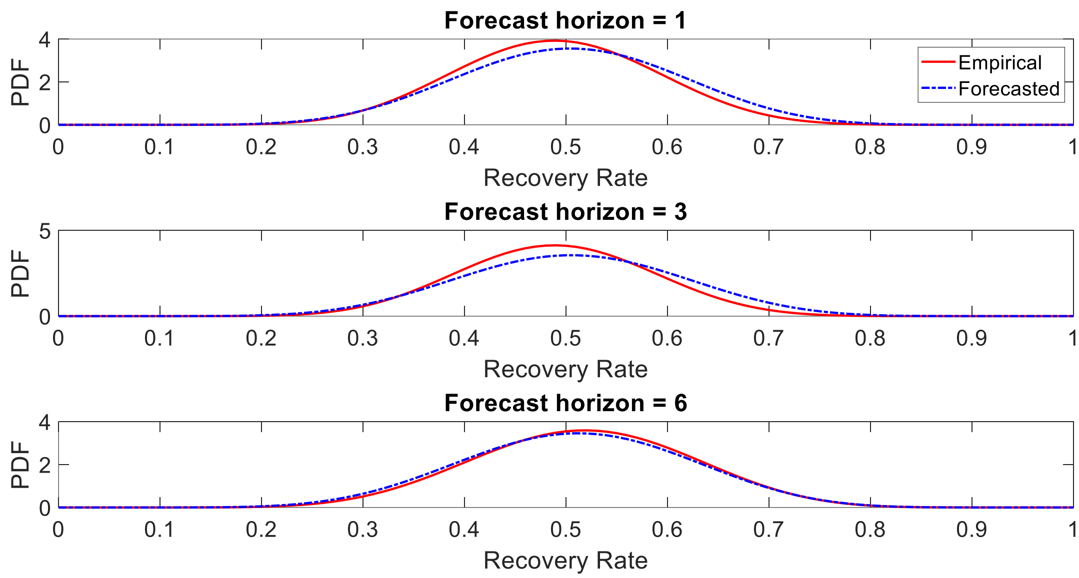


Figure 23 Empirical vs Forecast for 5-year CDS

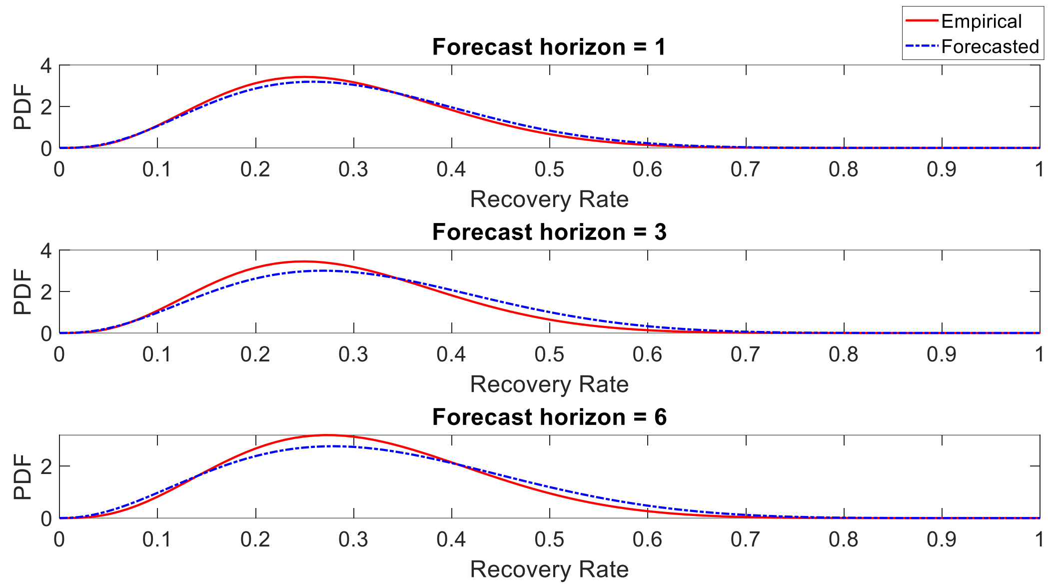


Figure 24 Empirical vs Forecast for 10-year CDS

Conclusion

In this paper, we study the distribution of market-implied recovery rates extracted from CDS data. The beta distribution modelling has been implemented among industries and the whole sample. It provides a good fit to the empirical recovery rate distribution. A bigger sample size helps to improve the fitting.

The market-implied recovery rates change with time as clearly reported by their mean before, during and after the recession. Therefore, it is critical to model them in the time series. I used a VAR (p) specification in level for the time series analysis. The quality of the in-sample fitting differs across industries. Similar patterns are observed in forecasting, too. One of the findings is that the benefit of sampling size could carry through from beta distribution estimation to the time series analysis and even to the forecasting ability. Also, the 6-month forecasting quality is satisfying.

Among the three maturities examined, the 5-year distribution obtain the best results. Forecasting the mean parameter yields much better results than forecasting the precision parameter.

To summarize, the results are encouraging. The paper demonstrated that we could model the forward-looking market-implied recovery rates with a simple beta distribution and a VAR model for its parameters. If a more sophisticated model or more factors are included in the modelling, it may further improve the quality for forecasting. This is important for the risk management or the pricing of credit-related instruments.

Bibliography

Acharya, V. V., Bharath, S. T. et Srinivasan, A. (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries, Journal of Financial Economics, vol. 85, no 3, p. 787-821.

Altman, E. I, Brady, B., Resti, A.et Sironi, A. (2005). The link between default and recovery rates: Theory, empirical evidence, and implications, Journal of Business, vol. 78, no 6, p. 2203-2228.

Altman, E. I. et Kishore, V. M. (1996). Almost everything you wanted to know about recoveries on defaulted bonds, Financial Analysts Journal, vol. 52, no 6, p. 57-64.

Bellotti, T. et Crook, J. (2012). Loss given default models incorporating macroeconomic variables for credit cards, International Journal of Forecasting, vol. 28, no 1, p. 171-182.

Berd, A.M. (2005). Recovery Swaps. Journal of Credit Risk, vol. 1(3), p. 61-70 (2005)

Bruche, M. et Aguado, C. G. (2010). Recovery rates, default probabilities, and the credit cycle, Journal of Banking and Finance, vol. 34, no 4, p. 754-764.

Chen, R., Zhou, H., Jin, C. et Zheng, W.(2019). Modeling of recovery rate for a given default by non-parametric method, Pacific-Basin Finance Journal, vol. 57, p. 101085.

Cox, J., C., Stephen, A. R. et Rubinstein, M.(1979). Option pricing: A simplified approach , Journal of Financial Economics, vol. 7, no 3, p. 229-263.

Das, S. R. et Hanouna, P. (2009). Implied recovery, Journal of Economic Dynamics and Control, vol. 33, no 11, p. 1837-1857.

Ferrari, S. et Neto, F.C.(2004). Beta regression for modelling rates and proportions, Journal of Applied Statistics, vol. 31, no 7, p. 799-815.

Finger, C., Finkelstein, V.,Lardy, J.-P.,Pan, G.,Ta, T.,Tierney, J.,2002.Credit Grades Technical Document, Risk Metrics.

François, P. et Jiang, W. (2019). Credit value adjustment with market-implied recovery, Journal of Financial Services Research, vol. 56, no 2, p. 145-166.

Gupton, G. et Stein R. (2005). Losscalc v2: Dynamic prediction of lgd - modeling methodology, Framework, p. 1-44.

Giesecke, K., Longstaff, F.A., Schaefer, S. et Strebulaev, I. (2011). Corporate bond default risk: A 150-year perspective, Journal of Financial Economics -Amsterdam-, vol. 102, no 2, p. 233-250.

Hurlin, C., Leymarie J. et Patin A.(2018). Loss functions for loss given default model comparison, European Journal of Operational Research, vol. 268, no 1, p. 348-360.

Kaposty, F., Kriebel J. et Löderbusch M.(2020). Predicting loss given default in leasing: A closer look at models and variable selection, International Journal of Forecasting, vol. 36, no 2, p. 248-266.

Khieu, H. D., Mullineaux, D. J. et Yi, H. C. (2012). The determinants of bank loan recovery rates, Journal of Banking and Finance, vol. 36, no 4, p. 923-933.

Lipton, A. et Sepp, A. (2009). Credit value adjustment for credit default swaps via the structural default model, The Journal of Credit Risk, vol. 5, no 2, p. 127-150.

Ljung, G. M. et Box, G. E. P. (1978). On a measure of lack of fit in time series models, Biometrika, vol. 65, no 2, p. 297-303.

Matuszyk, A., Mues, C. et Thomas, L. C. (2010). Modelling lgd for unsecured personal loans: Decision tree approach, The Journal of the Operational Research Society, vol. 61, no 3, p. 393-398.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates, The Journal of Finance, vol. 29, no 2, p. 449.

Nazemi, A., Heidenreich, K. et Fabozzi, F.J.(2018). Improving corporate bond recovery rate prediction using multi-factor support vector regressions, European Journal of Operational Research, vol. 271, no 2, p. 664-675.

Phillips, P. C. B. (1998). Impulse response and forecast error variance asymptotics in nonstationary vars, Journal of Econometrics, vol. 83, no 1-2, p. 21-56.

Qi, M.et Zhao, X.(2011). Comparison of modeling methods for loss given default, Journal of Banking and Finance, vol. 35, no 11, p. 2842-2855.

Sabato, G., & Schmid, M. (2008). Estimating Conservative Loss Given Default. Monetary Economics.

Schwert, G. W. (1989). Tests for unit roots: A monte carlo investigation, Journal of Business & Economic Statistics, vol. 7, no 2, p. 147-159.

Shleifer, Andrei et Robert W. Vishny (1992). Liquidation values and debt capacity: A market equilibrium approach, The Journal of Finance, vol. 47, no 4, p. 1343-1366.

Sims, C. A., Stock, J. H. et Watson, M. W. (1990). Inference in linear time series models with some unit roots, Econometrica, vol. 58, no 1, p. 113-144.

Tobback, E., Martens, D., Gestel, T. V. et Baesens, B. (2014). Forecasting loss given default models: Impact of account characteristics and the macroeconomic state, Journal of Operational Research Society, vol. 65, no 3, p. 376-392.

Unal, H., Madan D.et Guntay L.(2003). Pricing the risk of recovery in default with absolute priority rule violation , Journal of Banking & Finance., vol. 27, no 6, p. 1001.

Varma, P. et antor R.(2005). Determinants of recovery rates on defaulted bonds and loans for north american corporate issuers: 1983-2003, Journal of Fixed Income, vol. 14, no 4, p. 29-44.

Yao, X., Crook J. et Andreeva G. (2015). Support vector regression for loss given default modelling, European Journal of Operational Research, vol. 240, no 2, p. 528-538.

Appendix

A.1 Alpha and beta estimation using moment-matching method

Statistics for *μ*

	Manufacturing (134)	Transportation (64)	Finance (57)	Retail Trade (28)	Mining (25)	Services (24)	Whole sample (351)
1-year horizon							
Max	9.18	0.96	0.96	0.96	0.95	0.95	0.95
Min	0.66	0.63	0.51	0.59	0.48	0.49	0.62
Mean	0.90	0.89	0.89	0.89	0.88	0.86	0.89
Std Dev.	0.06	0.06	0.08	0.08	0.08	0.08	0.07
5-year horizon							
Max	0.67	0.67	0.67	0.72	0.65	0.70	0.67
Min	0.41	0.35	0.32	0.34	0.29	0.27	0.37
Mean	0.55	0.55	0.54	0.56	0.54	0.54	0.55
Std Dev.	0.06	0.06	0.07	0.07	0.06	0.07	0.06
10-year horizon							
Max	0.47	0.48	0.46	0.53	0.45	0.50	0.46
Min	0.26	0.22	0.20	0.22	0.19	0.17	0.24
Mean	0.34	0.35	0.34	0.36	0.34	0.34	0.34
Std Dev.	0.05	0.06	0.05	0.05	0.05	0.06	0.05

Statistics for *ω*

	Manufacturing (134)	Transportation (64)	Finance (57)	Retail Trade (28)	Mining (25)	Services (24)	Whole sample (351)
1-year horizon							
Max	123.97	219.56	258.27	317.57	417.21	204.91	104.01
Min	1.59	1.20	1.59	0.70	1.15	0.66	1.45
Mean	14.63	19.16	37.71	54.59	21.27	10.30	9.57
Std Dev.	17.90	39.60	48.13	69.22	52.23	25.73	12.88
5-year horizon							
Max	29.93	29.32	45.12	51.03	57.37	43.97	23.80
Min	5.94	4.41	4.83	3.20	3.84	3.04	5.18
Mean	16.82	15.24	21.10	16.89	18.16	14.97	15.76
Std Dev.	4.91	5.84	8.24	9.24	9.76	8.24	4.24
10-year horizon							
Max	23.16	24.54	27.36	36.32	56.97	40.17	19.49
Min	7.16	5.65	2.45	4.06	4.40	3.35	6.05
Mean	13.76	14.16	15.12	14.48	18.32	16.00	14.07
Std Dev.	3.19	3.73	4.86	5.79	8.45	7.09	3.10

A.2 VAR parameters for different industries

Manufacturing

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Constant(1)	0.370	6.914	0.245	5.793	0.102	3.071
Constant(2)	-14.304	-0.838	2.832	0.575	3.299	1.665
AR{1}(1,1)	0.559	8.521	0.334	4.731	0.407	5.188
AR{1}(2,1)	42.463	2.031	5.754	0.700	-2.398	-0.511
AR{1}(1,2)	0.000	1.108	-0.001	-2.236	0.001	0.529
AR{1}(2,2)	0.433	5.942	0.370	5.144	0.177	2.153
AR{2}(1,1)			0.242	3.501	0.320	4.107
AR{2}(2,1)			-4.279	-0.530	2.858	0.613
AR{2}(1,2)			0.000	0.278	0.000	-0.178
AR{2}(2,2)			0.280	3.836	0.261	3.232
AR{3}(1,1)			0.245	5.793	0.102	3.071
AR{3}(2,1)			2.832	0.575	3.299	1.665
AR{3}(1,2)			0.334	4.731	0.407	5.188
AR{3}(2,2)			5.754	0.700	-2.398	-0.511
AR{4}(1,1)			-0.001	-2.236	0.001	0.529

Transportation

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Constant(1)	0.265	3.390	0.158	4.411	0.224	5.157
Constant(2)	-59.992	-2.047	2.905	0.855	1.487	1.084
AR{1}(1,1)	0.364	4.709	0.485	6.655	0.331	4.084
AR{1}(2,1)	24.653	0.851	2.865	0.414	2.403	0.937
AR{1}(1,2)	0.000	-0.249	0.000	0.579	0.005	1.843
AR{1}(2,2)	0.388	5.037	0.451	6.308	0.178	2.259
AR{2}(1,1)	0.139	1.728	0.233	3.221	0.131	1.620
AR{2}(2,1)	52.852	1.753	-3.941	-0.573	-0.489	-0.191
AR{2}(1,2)	0.000	0.066	0.000	-0.243	-0.001	-0.577
AR{2}(2,2)	-0.059	-0.723	0.310	4.296	0.260	3.234
AR{3}(1,1)	0.246	3.115				
AR{3}(2,1)	7.643	0.258				
AR{3}(1,2)	0.000	1.323				
AR{3}(2,2)	-0.109	-1.349				
AR{4}(1,1)	-0.011	-0.137				
AR{4}(2,1)	12.450	0.411				
AR{4}(1,2)	0.000	-2.172				
AR{4}(2,2)	0.066	0.816				
AR{5}(1,1)	-0.053	-0.677				
AR{5}(2,1)	2.127	0.073				
AR{5}(1,2)	0.000	1.884				
AR{5}(2,2)	0.134	1.728				

Finance

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Constant(1)	0.515	8.512	0.235	4.930	0.305	8.851
Constant(2)	10.172	1.090	0.642	0.223	1.222	1.317
AR{1}(1,1)	0.383	5.223	0.401	5.507	0.273	3.644
AR{1}(2,1)	3.401	0.301	-0.478	-0.108	0.793	0.394
AR{1}(1,2)	0.000	-0.221	-0.003	-2.377	-0.004	-1.419
AR{1}(2,2)	0.319	4.313	0.634	8.461	0.522	7.640
AR{2}(1,1)			-0.002	-0.024		
AR{2}(2,1)			-0.031	-0.007		
AR{2}(1,2)			0.002	1.415		
AR{2}(2,2)			0.005	0.053		
AR{3}(1,1)			0.179	2.503		
AR{3}(2,1)			2.234	0.517		
AR{3}(1,2)			-0.001	-0.655		
AR{3}(2,2)			0.117	1.536		

Retail Trade

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Constant(1)	0.441	6.127	0.192	4.174	0.211	5.218
Constant(2)	-7.265	-0.424	4.986	1.854	1.642	3.081
AR{1}(1,1)	0.271	3.520	0.414	5.561	0.236	3.296
AR{1}(2,1)	18.955	1.035	-2.963	-0.680	0.327	0.347
AR{1}(1,2)	0.000	1.231	0.000	-0.019	0.006	1.146
AR{1}(2,2)	0.111	1.490	0.174	2.344	0.168	2.318
AR{2}(1,1)	0.179	2.334	0.101	1.260	0.275	3.886
AR{2}(2,1)	13.223	0.724	4.601	0.980	0.319	0.343
AR{2}(1,2)	0.000	-0.233	0.000	-0.279	-0.004	-0.795
AR{2}(2,2)	0.273	3.652	0.152	2.054	0.177	2.446
AR{3}(1,1)			0.157	2.125		
AR{3}(2,1)			-4.861	-1.125		
AR{3}(1,2)			0.001	0.676		
AR{3}(2,2)			0.180	2.450		

Mining

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Constant(1)	0.608	6.849	0.210	4.888	0.255	8.384
Constant(2)	-45.454	-0.805	2.672	0.440	0.700	0.763
AR{1}(1,1)	0.074	1.012	0.303	4.295	0.350	5.065
AR{1}(2,1)	44.461	0.951	5.402	0.541	5.400	2.589
AR{1}(1,2)	0.000	-1.590	-0.001	-1.558	0.007	2.832
AR{1}(2,2)	0.074	0.987	0.040	0.533	0.270	3.750
AR{2}(1,1)	0.248	3.388	0.346	4.937		
AR{2}(2,1)	34.101	0.732	1.391	0.141		
AR{2}(1,2)	0.000	0.374	0.000	0.578		
AR{2}(2,2)	0.175	2.334	0.012	0.156		

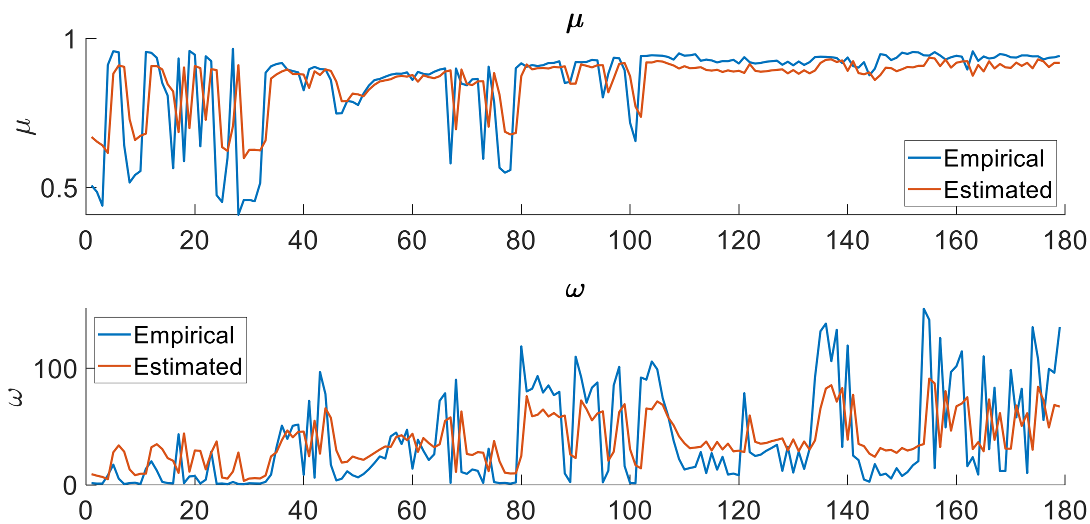
Services

VAR Parameter	1-year CDS		5-year CDS		10-year CDS	
	Value	*t*-stat	Value	*t*-stat	Value	*t*-stat
Constant(1)	0.284	4.595	0.202	4.158	0.152	4.407
Constant(2)	-17.297	-0.438	-1.259	-0.424	2.422	2.332
AR{1}(1,1)	0.426	5.625	0.333	4.435	0.393	5.320
AR{1}(2,1)	66.151	1.363	2.101	0.458	-2.302	-1.032
AR{1}(1,2)	0.000	0.410	0.000	0.034	0.002	0.947
AR{1}(2,2)	0.079	1.009	0.361	4.917	0.158	2.116
AR{2}(1,1)	0.255	3.342	0.241	3.047	0.211	2.861
AR{2}(2,1)	-0.628	-0.013	5.587	1.157	3.867	1.735
AR{2}(1,2)	0.000	-0.664	0.001	0.756	0.000	-0.042
AR{2}(2,2)	0.033	0.432	0.122	1.561	0.047	0.632
AR{3}(1,1)			0.025	0.312		
AR{3}(2,1)			-0.921	-0.191		
AR{3}(1,2)			-0.001	-0.428		
AR{3}(2,2)			0.077	0.978		
AR{4}(1,1)			0.060	0.812		
AR{4}(2,1)			-2.014	-0.444		
AR{4}(1,2)			-0.001	-0.938		
AR{4}(2,2)			0.229	3.090		

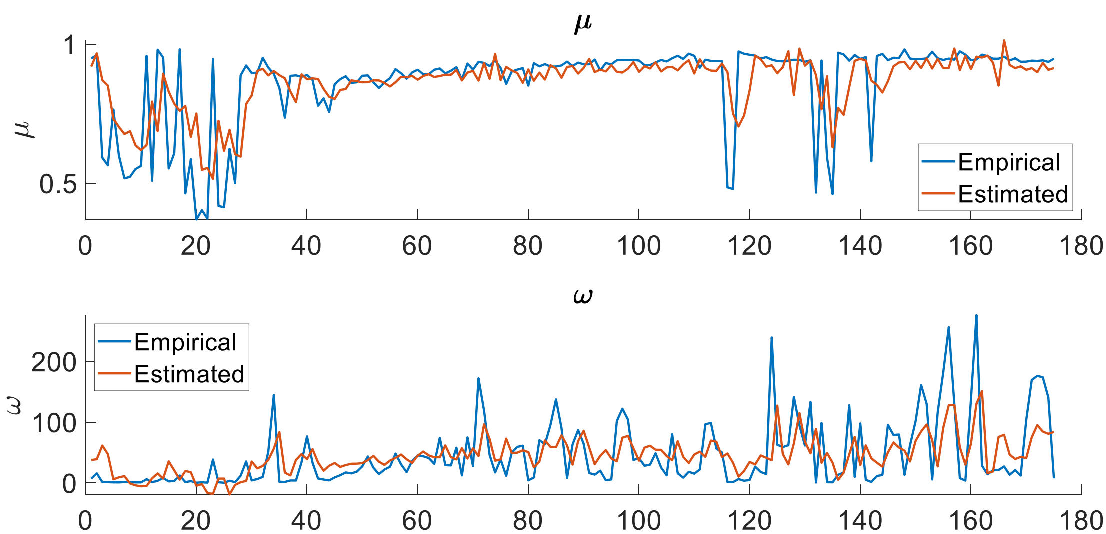
A.3 Empirical vs Estimated for different industries

1-year CDS

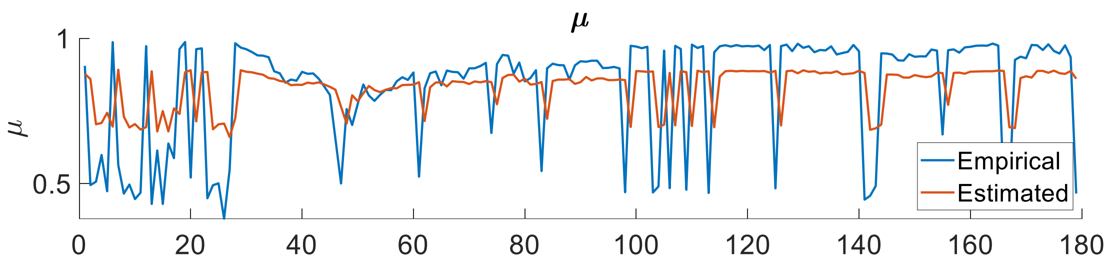
Manufacturing

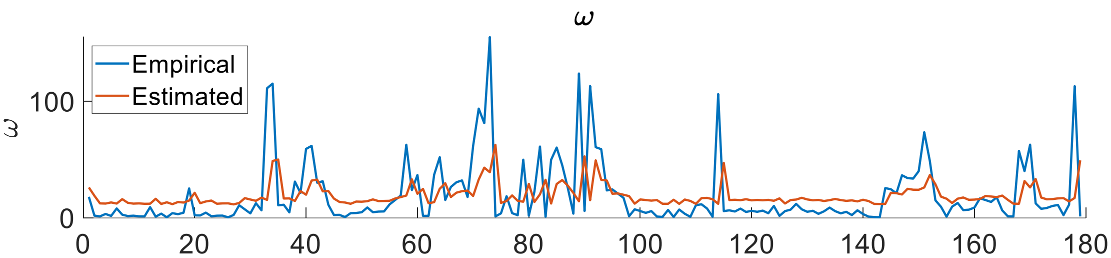


Transportation

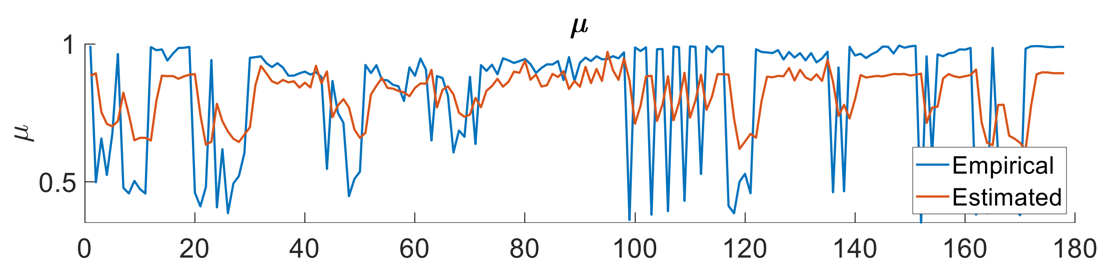


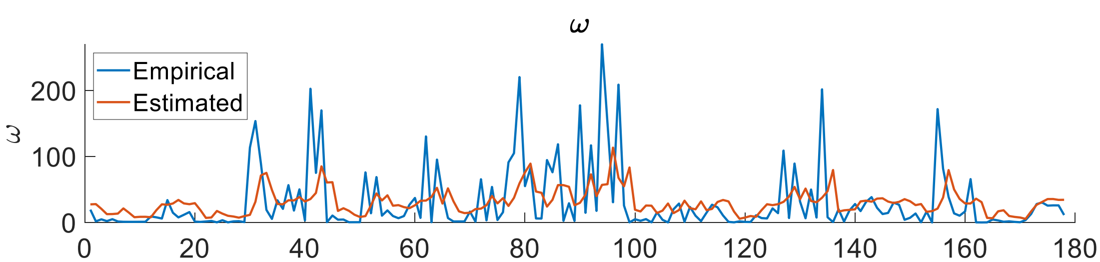
Finance



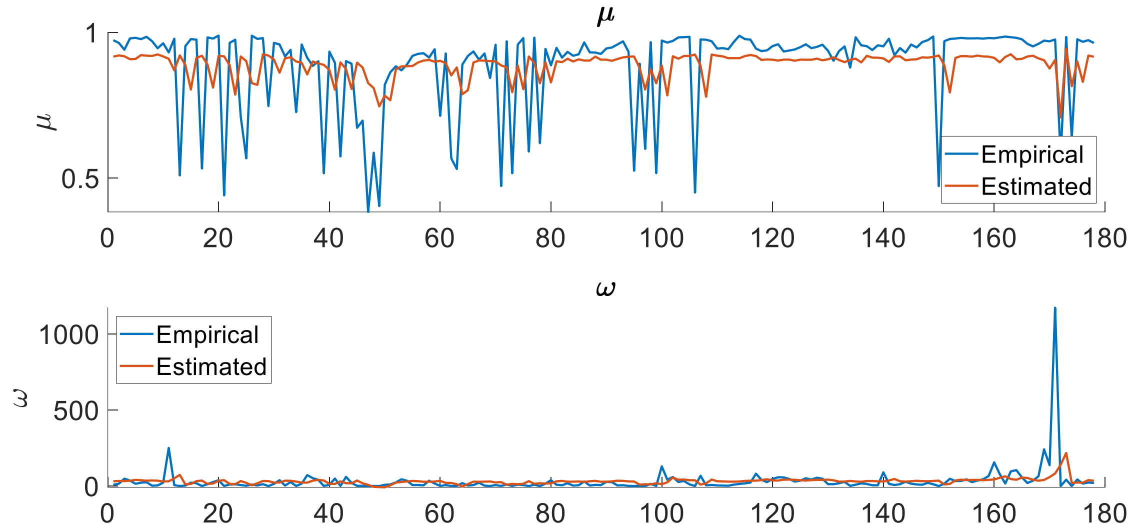


Retail Trade

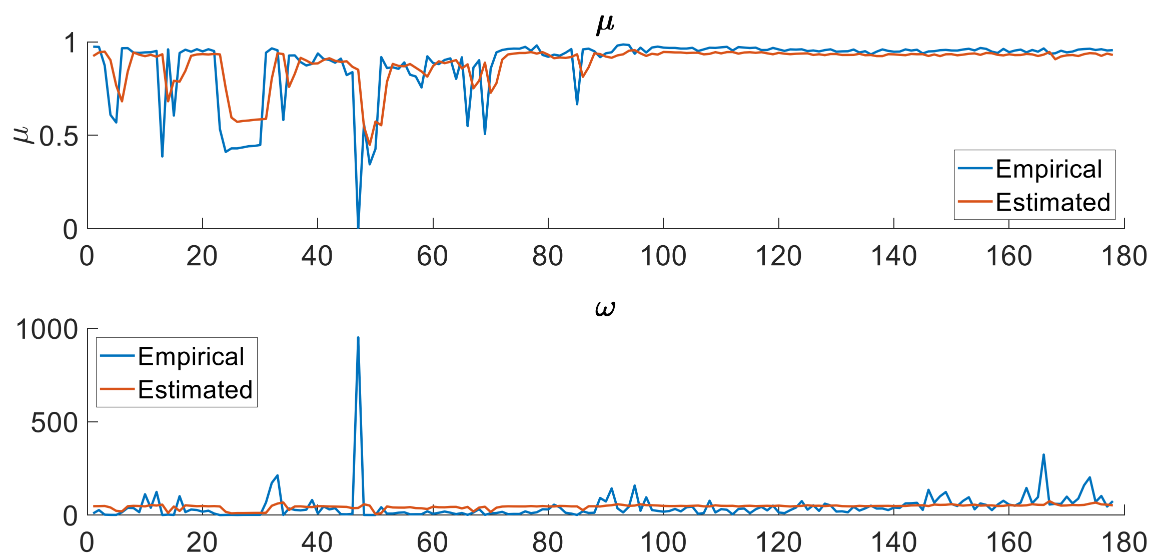




Mining

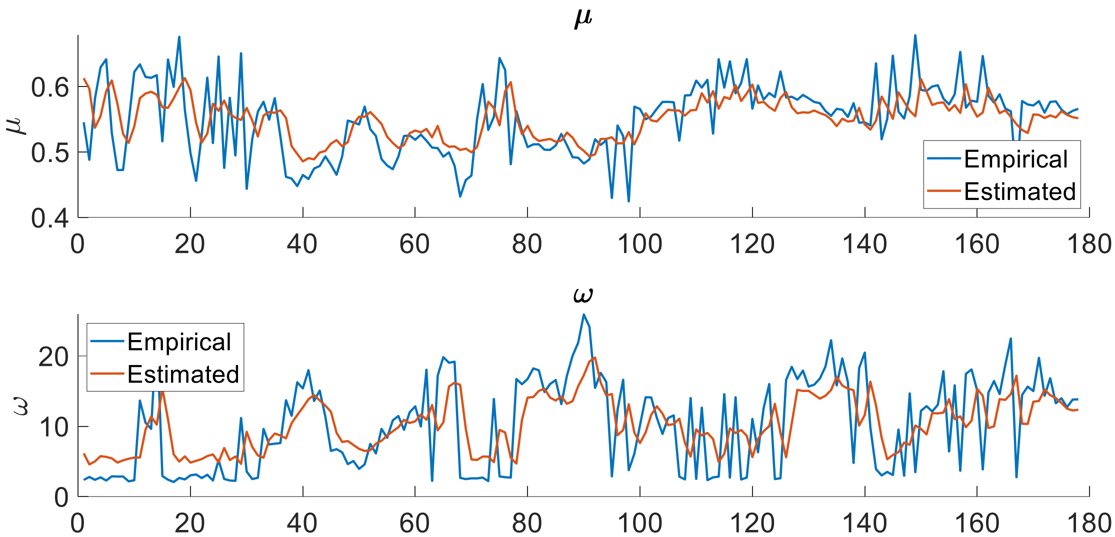


Services

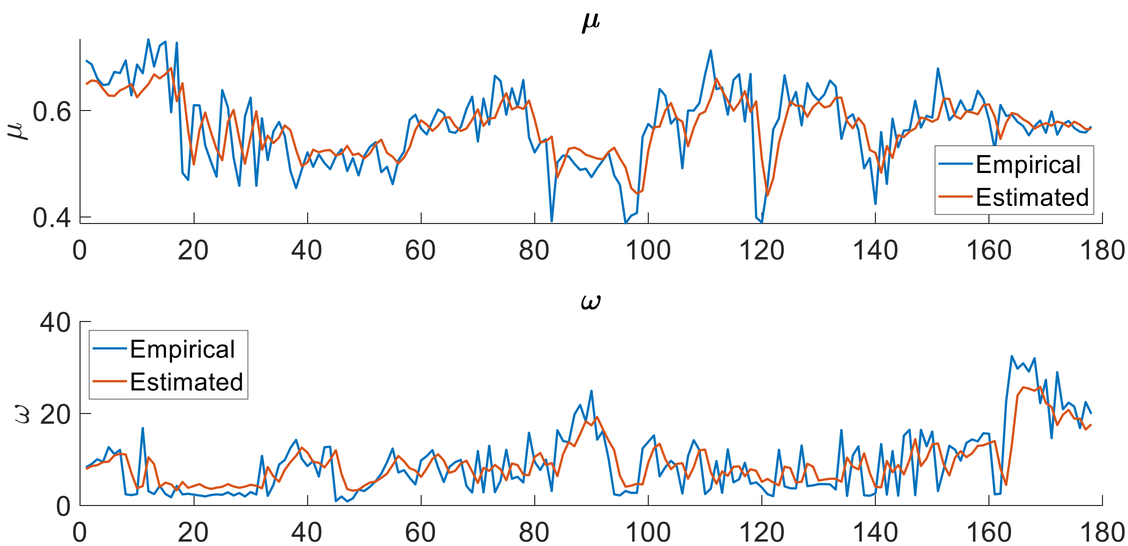


5-year CDS

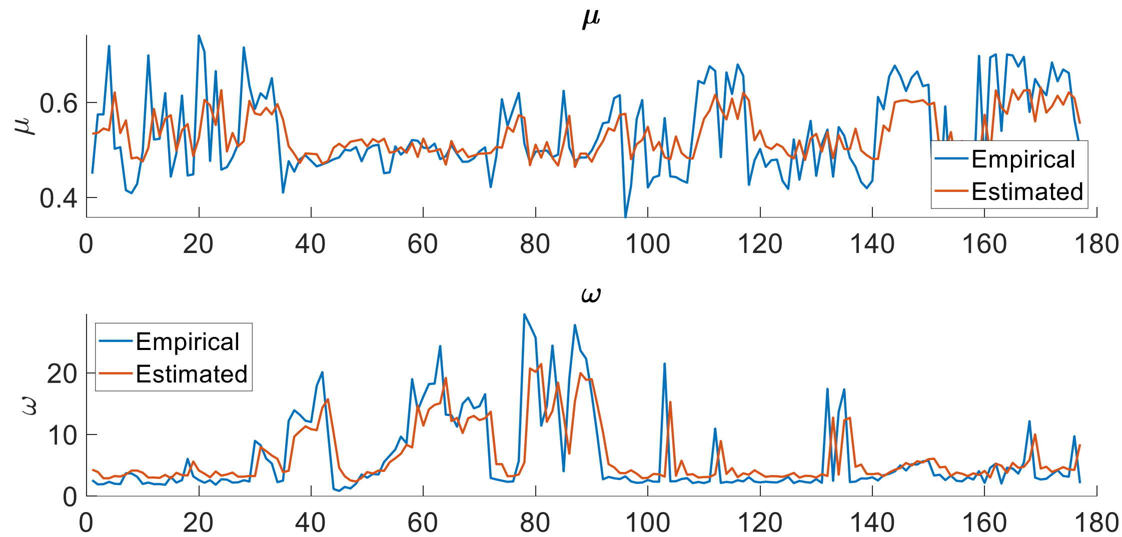
Manufacturing



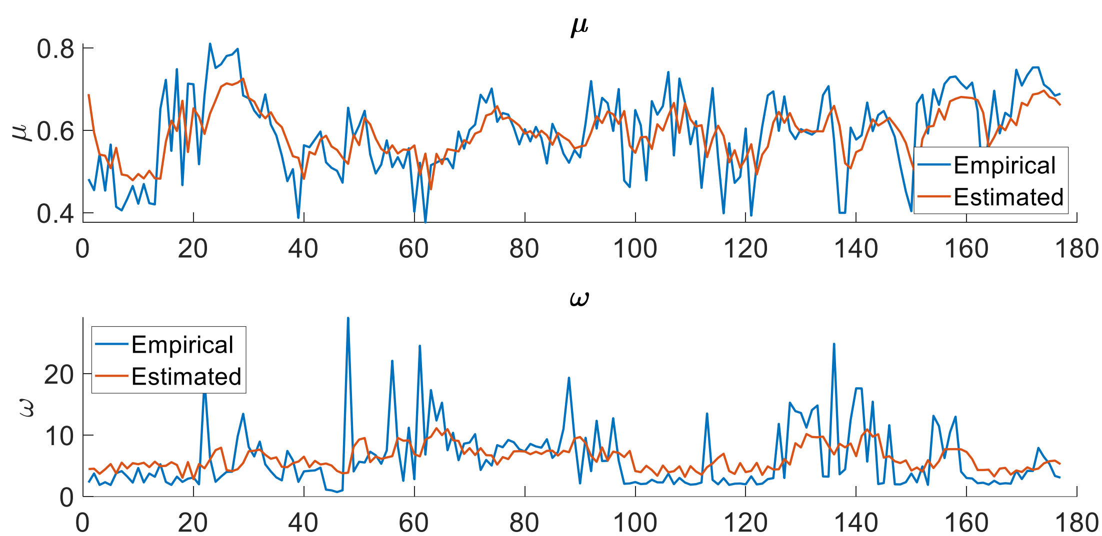
Transportation



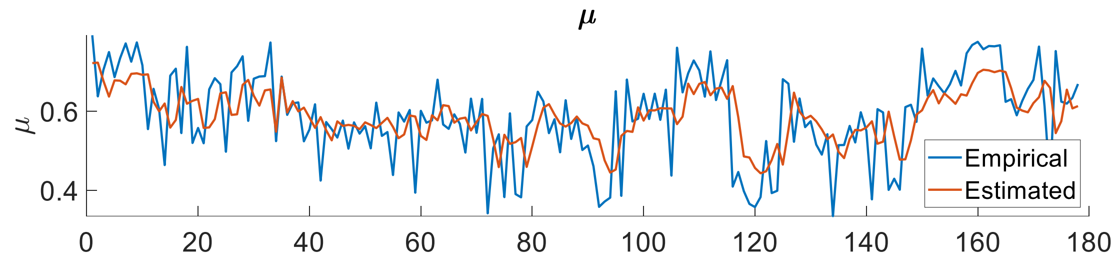
Finance

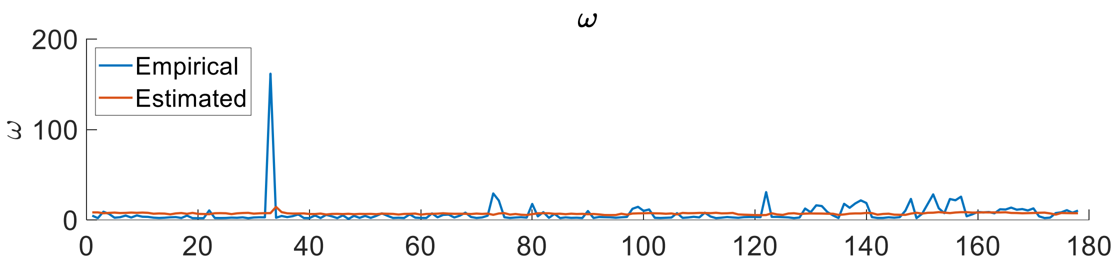


Retail Trade

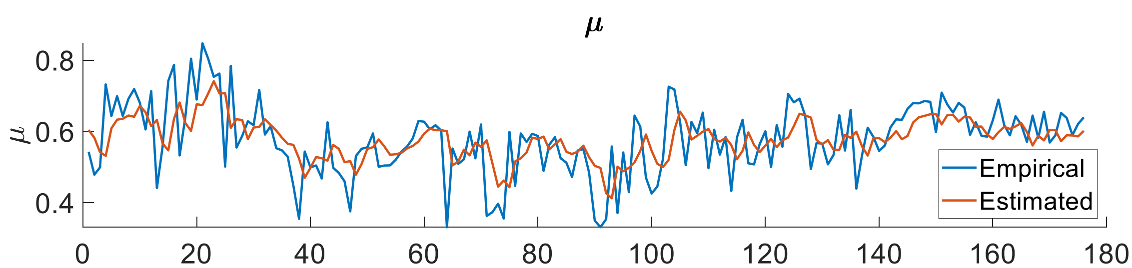


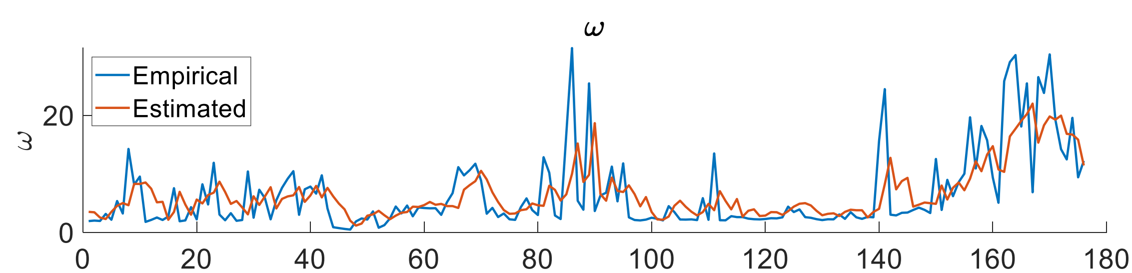
Mining





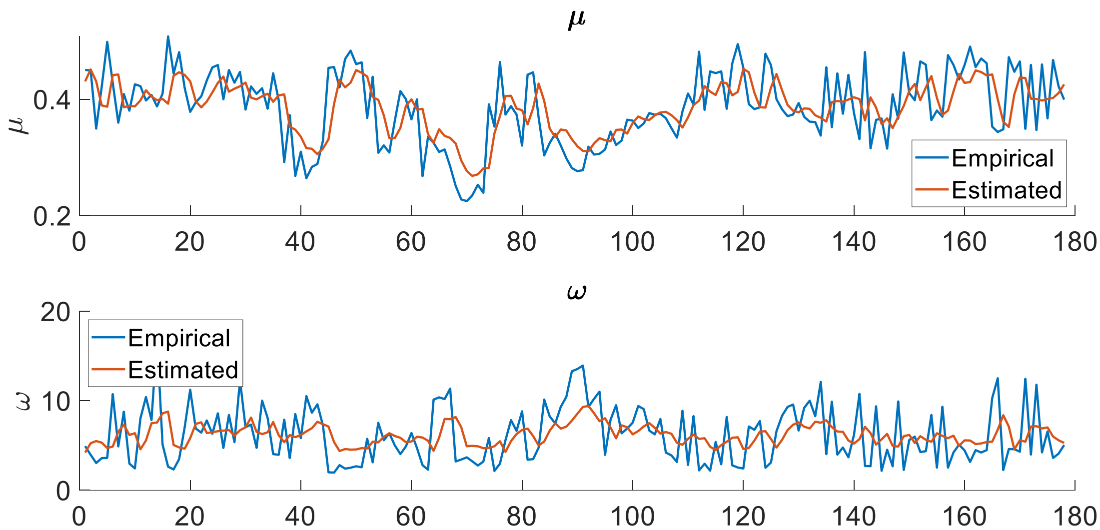
Services



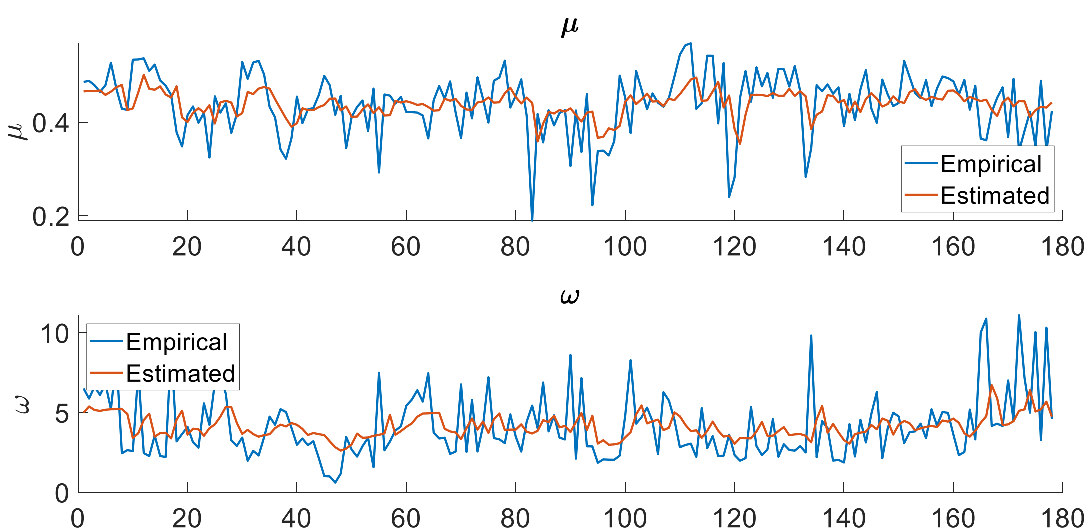


10-year CDS

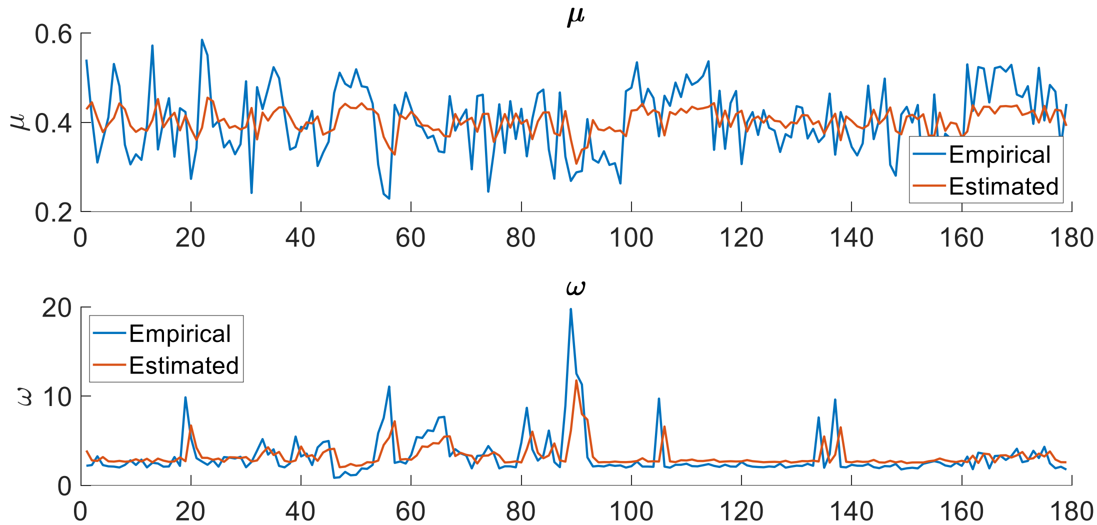
Manufacturing



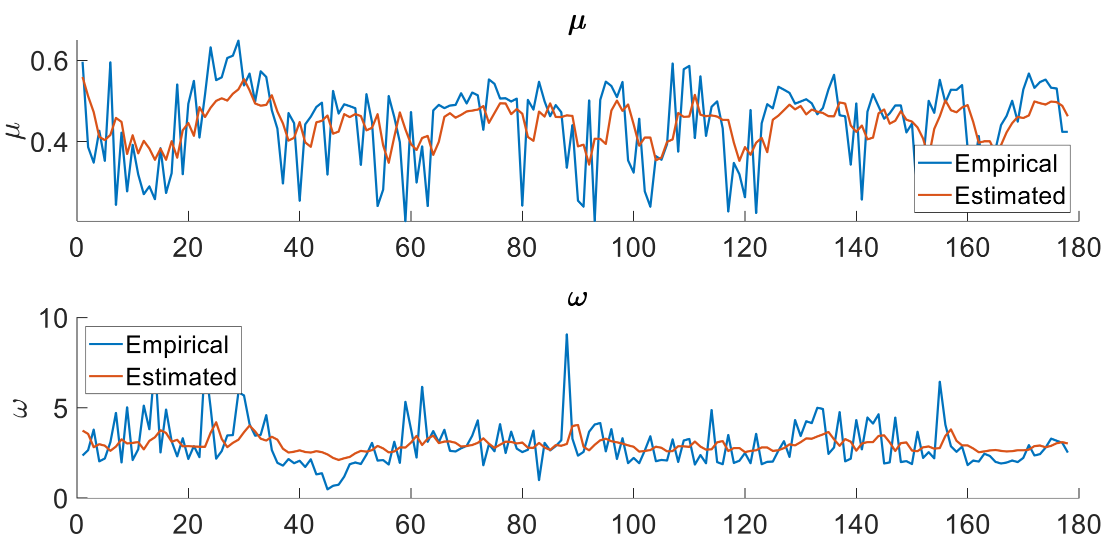
Transportation



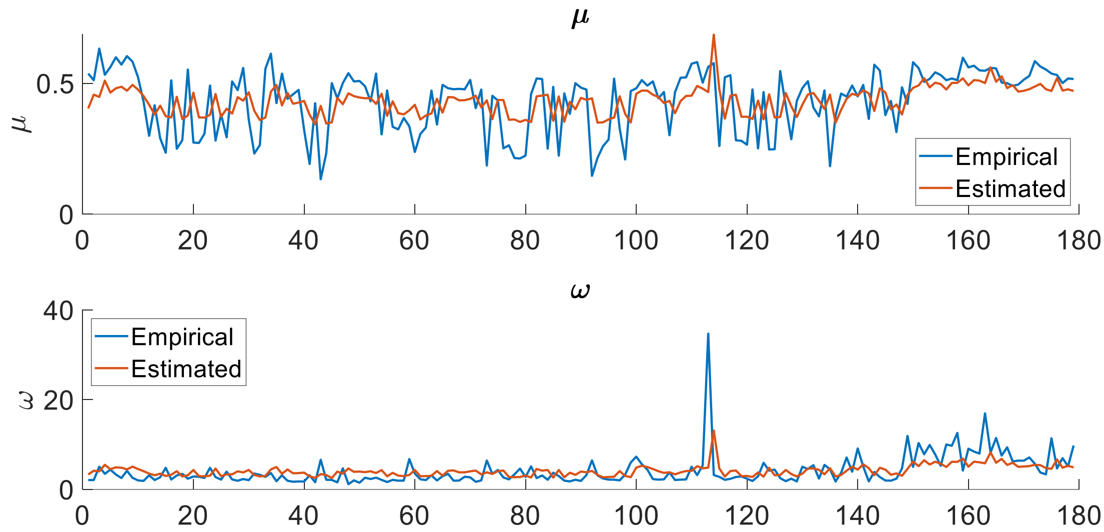
Finance



Retail Trade



Mining



Services